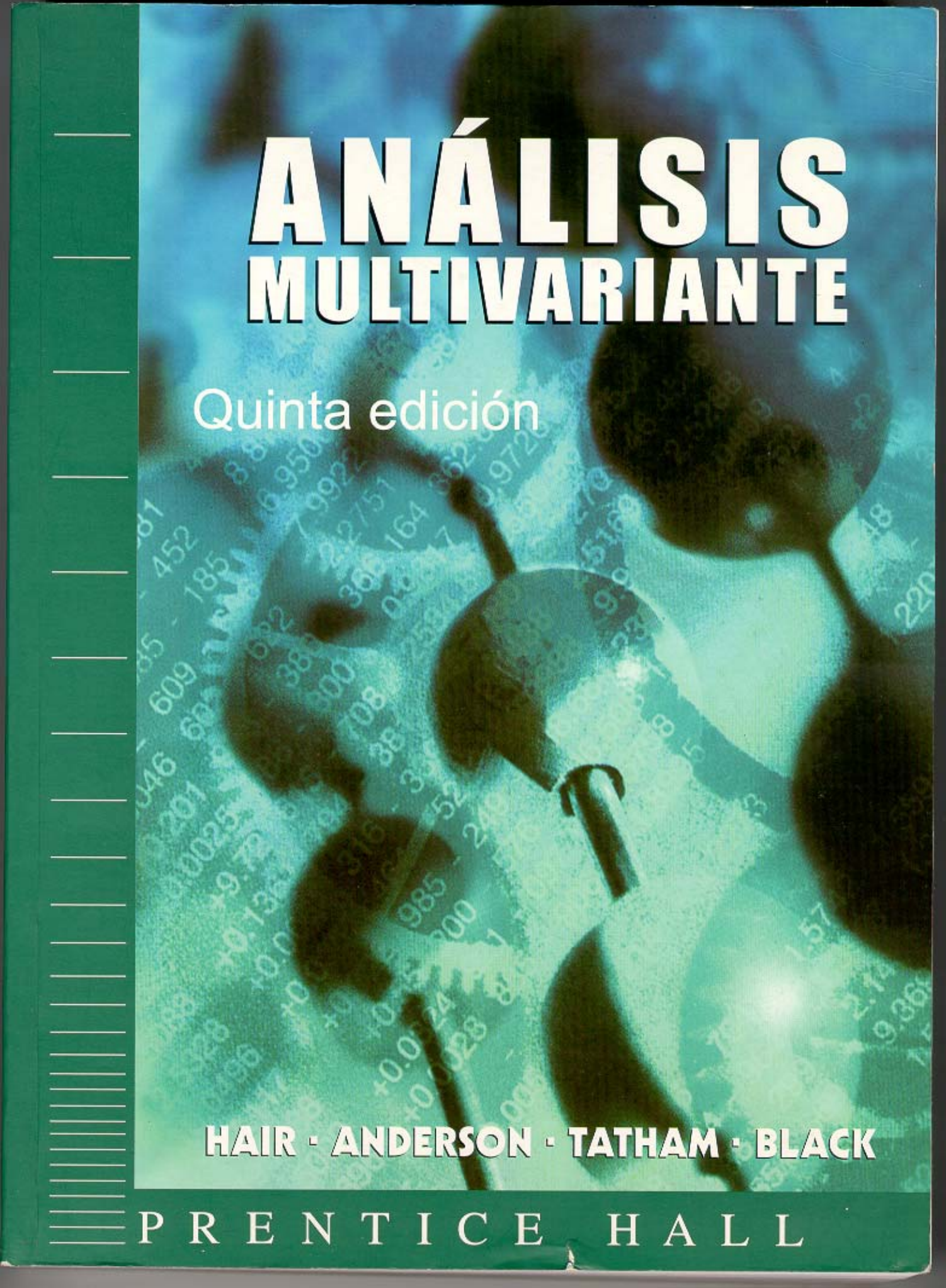

ANÁLISIS
MULTIVARIANTE
Quinta edición
HAIR · ANDERSON · TATHAM · BLACK
PRENTICE HALL

ANÁLISIS MULTIVARIANTE

Quinta edición

Joseph F. HAIR, Jr.
Louisiana State University

Rolph E. ANDERSON
Drexel University

Ronald L. TATHAM
Burke Marketing Research

William C. BLACK
Louisiana State University

Traducción:

Esme Prentice

Diego Cano
Universidad Autónoma de Madrid

Revisión técnica y compilación de las lecturas complementarias:

Mónica Gómez Suárez
Profesora titular de Comercialización e Investigación de Mercados
Universidad Autónoma de Madrid

PRENTICE HALL
Madrid • Upper Saddle River • Buenos Aires • Londres • México • Nueva Delhi • Río de Janeiro
Santafé de Bogotá • Singapur • Sydney • Tokio • Toronto

Datos de catalogación bibliográfica

J. F. Hair, Jr., R. E. Anderson, R. L. Tatham, W. C. Black
ANÁLISIS MULTIVARIANTE, 5.ª ed.
PRENTICE HALL IBERIA, Madrid, 1999

ISBN: 84-8322-035-0
Materia: Estadística-Teoría, procedimientos y métodos. 311

Formato 195 × 265 Páginas: 832

J. F. Hair, Jr., R. E. Anderson. R. L. Tatham, W. C. Black
ANÁLISIS MULTIVARIANTE, 5.ª ed.

No está permitida la reproducción total o parcial de esta obra ni su tratamiento o transmisión por cualquier medio o método sin autorización escrita de la Editorial.

DERECHOS RESERVADOS
© 1999 respecto a la primera edición en español por:
PRENTICE HALL IBERIA, S.R.L.

ISBN: 84-8322-035-0
Depósito Legal: M 31.513-1999

Traducido de:
MULTIVARIATE DATA ANALYSIS, Fifth edition
Prentice Hall International, Inc.
© MCMXCVIII
ISBN: 0-13-930587-4

Edición en español:
Editor: Andrés Otero
Asistente editorial: Ana Isabel García
Diseño de cubierta: Domingo Rufo
Composición: COPIBOOK, S. L.
Impreso por:

IMPRESO EN ESPAÑA - PRINTED IN SPAIN

Este libro ha sido impreso con papel y tintas ecológicos

Contenido

Prólogo XVII

Prólogo a la edición en castellano XXI

Nota introductoria XXIII

Profesores que han participado con lecturas complementarias XXV

Capítulo 1. Introducción 1

¿Qué es el análisis multivariante? 2
El impacto de la revolución informática 3
Definición del análisis multivariante 4
Algunos conceptos básicos del análisis multivariante 4
El valor teórico 4
Escalas de medida 5
Error de medida y medidas multivariantes 6
Significación estadística frente a potencia estadística 8
Tipos de técnicas multivariantes 11
Componentes principales y análisis factorial común 11
Regresión múltiple 11
Análisis discriminante múltiple 11
Análisis multivariante de la varianza y covarianza 12
Análisis conjunto 12
Correlación canónica 12
Análisis cluster 13
Análisis multidimensional 13
Análisis de correspondencias 13
Modelos de probabilidad lineal 14
Modelos de ecuaciones estructurales 14
Otras técnicas multivariantes emergentes 14
Una clasificación de las técnicas multivariantes 15
Orientaciones para el análisis multivariante y su interpretación 19
Establecer la significación práctica así como la estadística 19
El tamaño muestral afecta a todos los resultados 20
Conocer los datos 20
Procurar la parsimonia del modelo 20
Atender a los errores 21
Validar los resultados 21
Una aproximación organizada al diseño de modelos multivariantes 21
Primer paso: Definir el problema de investigación, objetivos y técnica multivariante conveniente 22
Segundo paso: Desarrollo del proyecto de análisis 22
Tercer paso: Evaluación de los supuestos básicos de la técnica multivariante 23
Cuarto paso: Estimación del modelo multivariante y valoración del ajuste del modelo 23

Quinto paso: Interpretar el valor teórico 23
Sexto paso: Validación del modelo multivariante 24
Un diagrama de flujos de decisiones 24
Bases de datos 24
La principal base de datos 24
Otras bases de datos 26
Organización de los capítulos restantes 26
Resumen 27
Cuestiones de revisión 27
Bibliografía 28

Sección 1. Preparación para el análisis multivariante 29

Capítulo 2. Análisis previo de los datos 31

Introducción 32
Examen gráfico de los datos 33
La naturaleza de la variable: Análisis de la forma de la distribución 33
Análisis de relación entre variables 35
Análisis de las diferencias entre grupos 35
Perfiles multivariantes 37
Resumen 39
Datos ausentes 39
Un ejemplo sencillo del análisis de datos ausentes 40
Conocimiento de las razones causantes de datos ausentes 41
Examen de los tipos de datos ausentes 42
Diagnóstico de la aleatoriedad en el proceso de pérdida de observaciones 43
Aproximaciones al tratamiento de datos ausentes 43
Utilizar sólo aquellas observaciones con datos completos 44
Supresión de caso(s) y/o variable(s) 44
Métodos de imputación 44
Procedimientos basados en el modelo 47
Un ejemplo de diagnóstico de datos ausentes 48
Un repaso del análisis de valores ausentes 55
Resumen 56
Casos atípicos 57
Detección de casos atípicos 57
Descripción de casos atípicos y especificación 59
Mantenimiento o eliminación de los casos atípicos 59
Un ejemplo del análisis de casos atípicos 59
Verificación de los supuestos del análisis multivariante 61
Valoración de las variables individuales frente al modelo univariante 63
Normalidad 63
Homocedasticidad 66
Linealidad 67
Ausencia de errores correlacionados 68
Transfornaciones de los datos 69
Un ejemplo de contrastación de los supuestos básicos del análisis multivariante 70
Homocedasticidad 74
La incorporación de datos no métricos con variable ficticias 75
Resumen 77
Cuestiones de revisión 77
Bibliografía 77

Capítulo 3. Análisis factorial 79

¿Qué es el análisis factorial? 80
Un ejemplo hipotético del análisis factorial 81
El proceso de decisión del análisis factorial 83

Primer paso: Objetivos del análisis factorial 83
Segundo paso: El diseño de un análisis factorial 86
Tercer paso: Supuestos en el análisis factorial 88
Cuarto paso: La estimación de los factores y la valoración del ajuste general 89
Quinto paso: Interpretación de los factores 94
Sexto paso: Validación del análisis factorial 102
Séptimo paso: Usos adicionales de los resultados del análisis factorial 103

Un ejemplo ilustrativo 107

Primer paso: Objetivos del análisis factorial 108
Segundo paso: Diseño del análisis factorial 108
Tercer paso: Supuestos del análisis factorial 108
Análisis de factores componentes: Del paso cuarto al séptimo 109
Análisis factorial común: Pasos cuarto y quinto 118
Una consideración gerencial de los resultados 121

Resumen 121
Cuestiones de revisión 122
Bibliografía 122
Lectura complementaria: *La imagen de la empresa detallista y su percepción por los consumidores: Una aplicación del análisis factorial*, por Rodolfo Vázquez Casielles 124

Sección 2. Técnicas de dependencia 141

Capítulo 4. Análisis de regresión múltiple 143

¿Qué es el análisis de regresión múltiple? 144
Un ejemplo de la regresión simple y multiple 145

Fijando una línea de base: predicción sin variable independiente 145
Predicción mediante una única variable independiente. Regresión simple 146
Predicción utilizando varias variables independientes: Análisis de regresión múltiple 151
Resumen 153

Un proceso de decisión para el análisis de regresión múltiple 153
Primer paso: Objetivos de la regresión múltiple 155

Problemas de investigación adecuados para la regresión múltiple 155
Especificación de la relación estadística 156
Selección de variables dependientes e independientes 157

Segundo paso: Diseño en la investigación del análisis de la regresión múltiple 158

Tamaño muestral 159
Predictores de efectos fijos frente a predictos de efectos aleatorios 160
Creación de variables adicionales 161

Tercer paso: Supuestos en el análisis de regresión múltiple 165

Valoración de las variables individuales frente al valor teórico 166
Linealidad del fenómeno 167
Varianza constante del término de error 168
Independencia de los términos de error 168
Normalidad de la distribución del término de error 169
Resumen 169

Cuarto paso: Estimación del modelo de regresión y valoración 169

Aproximaciones generales a la selección de variables 171
Contrastación del cumplimiento de los supuestos de regresión 174
Examen de la significación estadística de nuestro modelo 174

Identificación de observaciones influyentes ... 177
Quinto paso: Interpretación del valor teórico de la regresión ... 179
Utilización de los coeficientes de regresión ... 180
Estandarización de los coeficientes de regresión: Los coeficientes beta ... 180
Evaluación de la multicolinealidad ... 181
Sexto paso: Validación de los resultados ... 186
Muestras adicionales o muestras divididas ... 186
Cálculo del estadístico PRESS ... 186
Comparación de los modelos de regresión ... 187
Predicción con el modelo ... 187
Ejemplo de un análisis de regresión ... 187
Primer paso: Objetivos de la regresión múltiple ... 187
Segundo paso: Diseño de la investigación mediante el análisis de regresión múltiple ... 188
Tercer paso: Supuestos del análisis de regresión múltiple ... 188
Cuarto paso: Estimación del modelo de regresión y valoración global del ajuste ... 189
Quinto paso: Interpretación del resultado teórico ... 198
Sexto paso: Validación de los resultados ... 199
Evaluación de modelos de regresión alternativos ... 200
Una visión gerencial de los resultados ... 202
Resumen ... 204
Cuestiones de revisión ... 204
Bibliografía ... 204

Apéndice 4A. Validación avanzada en el análisis de regresión múltiple ... 207

Evaluación de la multicolinealidad ... 208
Un proceso en dos partes ... 209
Ejemplo de evaluación de la multicolinealidad ... 209
Identificación de observaciones influyentes ... 209
Primer paso: Examen de los residuos ... 210
Segundo paso: Identificación de punto de apalancamiento mediante los predictores ... 211
Tercer paso: Diagnóstico de un único caso de observaciones influyentes ... 212
Cuarto paso: Selección y tratamiento de observaciones influyentes ... 213
Ejemplo de la base de datos HATCO ... 213
Visión panorámica ... 224
Resumen ... 225
Cuestiones de revisión ... 225
Bibliografía ... 225
Lecturas complementarias: *Relación entre concentración y precios en el comercio minorista*, por M.ª Jesús Yagüe Guillén y *El comportamiento de búsqueda de precios de los consumidores: Un análisis en el mercado minorista de alimentación*, por Carmen Berné Manero, Marta Pedraja Iglesias y Pilar Rivera Torres ... 226

Capítulo 5. Análisis discriminante múltiple y regresión logística ... 249

¿Qué son el análisis discriminante y la regresión logística? ... 250
Analogía con la regresión y MANOVA ... 252
Un ejemplo hipotético del análisis discriminante ... 252
Análisis discriminante de dos grupos: Compradores frente a no compradores ... 253
Una representación geométrica de la función discriminante de dos grupos ... 256
Un ejemplo de análisis discriminante de tres grupos: Propósitos de cambio ... 257
El proceso de decisión en el análisis discriminante ... 260
Primer paso: Objetivos del análisis discriminante ... 261
Segundo paso: Diseño de la investigación mediante análisis discriminante ... 261
Selección de las variables dependiente e independientes ... 262
Tamaño muestral ... 262
División de la muestra ... 263

Tercer paso: Supuestos del análisis discriminante ... 264
Cuarto paso: Estimación del modelo discriminante y valoración del ajuste global ... 264
Método de cálculo ... 265
Significación estadística ... 266
Valoración del ajuste global ... 267
Diagnosis mediante casos ... 274
Resumen ... 275
Quinto paso: Interpretación de los resultados ... 275
Ponderaciones discriminantes ... 276
Cargas discriminantes ... 276
Valores parciales de la F ... 276
Interpretación de dos o más funciones ... 276
¿Cuál es el método de interpretación a emplear? ... 278
Sexto paso: Validación de los resultados ... 278
División de la muestra o procedimientos de validación cruzada ... 279
Perfilar las diferencias entre los grupos ... 279
Regresión logística: Regresión con una variable dependiente binaria ... 280
Representación de la variable dependiente binaria ... 280
Estimación del modelo de regresión logística ... 281
Interpretación de los coeficientes ... 281
Un ejemplo ilustrativo de dos grupos ... 284
Primer paso: Objetivos del análisis discriminante ... 284
Segundo paso: Diseño de la investigación mediante el análisis discriminante ... 285
Tercer paso: Supuestos del análisis discriminante ... 286
Cuarto paso: Estimación de la función discriminante y valoración del ajuste global ... 286
Quinto paso: Interpretación de los resultados ... 294
Sexto paso: Validación de los resultados ... 297
Una visión gerencial de conjunto ... 298
Un ejemplo ilustrativo de tres grupos ... 298
Primer paso: Objetivos del análisis discriminante ... 298
Segundo paso: Diseño de la investigación del análisis discriminante ... 299
Tercer paso: Supuestos del análisis discriminante ... 299
Cuarto paso: Estimación de la función discriminante y valoración del ajuste global ... 299
Quinto paso: Interpretación de los resultados del análisis discriminante de tres grupos ... 311
Sexto paso: Validación de los resultados discriminantes ... 315
Una visión gerencial de conjunto ... 315
Un ejemplo ilustrativo de regresión logística ... 316
Pasos primero, segundo y tercero: Diseño de objetivos, diseño de la investigación y supuestos estadísticos. 316
Cuarto paso: Estimación del modelo de regresión logística y valoración del ajuste global ... 316
Quinto paso: Interpretación de los resultados ... 322
Sexto paso: Validación de los resultados ... 322
Una visión gerencial de conjunto ... 322
Resumen ... 322
Cuestiones de revisión ... 323
Bibliografía ... 323
Lecturas complementarias: *Factores determinantes del nivel de equipamiento doméstico. Un contraste empírico*, por Carlos Flavián Blanco, Eva Martínez Salinas y Yolanda Polo Redondo y *Las estrategias de captación y retención de clientes y la lealtad. Un análisis de segmentación*, por Sandra Cavero Brújula y Javier Cebollada Calvo ... 325

Capítulo 6. Análisis multivariante de la varianza ... 345
¿Qué es el análisis multivariante de la varianza? ... 346
Procedimientos univariantes para valorar las diferencias de grupo ... 346
Análisis multivariante de la varianza (MANOVA) ... 349

Las diferencias entre MANOVA y el análisis discriminante 351
Una ilustración hipotética del MANOVA 352
¿Cuándo se debe utilizar MANOVA? 353
Control del porcentaje de errores experimentales 354
Diferencias entre una combinación de variables dependientes 354
Un proceso de decisión para el MANOVA 354
Primer paso: Objetivos del MANOVA 355
Tipos de preguntas adecuadas en el análisis multivariante del MANOVA 356
Segundo paso: Cuestiones básicas del diseño de la investigación mediante MANOVA 357
Requisitos de tamaño muestral — en conjunto y por grupos 357
Diseños factoriales — dos o más tratamientos 357
Uso de covariaciones— ANCOVA y MANCOVA 359
Tercer paso: Supuestos básicos del ANOVA y del MANOVA 362
Independencia 362
Igualdad de las matrices de varianzas-covarianzas 363
Normalidad 363
Linealidad y multicolinealidad entre las variables dependientes 364
Sensibilidad a los datos atípicos 364
Cuarto paso: Estimación del modelo MANOVA y valoración del ajuste global 364
Criterios para la contrastación de la significación 364
Potencia estadística de los contrastes multivariantes 366
Quinto paso: Interpretación de los resultados del MANOVA 368
Evaluación de las covarianzas 369
Evaluación del valor teórico dependiente 369
Identificación de las diferencias entre los distintos grupos 370
Sexto paso: Validación de los resultados 372
Resumen 372
Ejemplo 1: Diferencia entre dos grupos independientes 372
Un enfoque univariante: El contraste *t* 373
Un enfoque multivariante: La T^2 de Hotelling 375
Ejemplo 2: Diferencia entre *k* grupos independientes 380
Un enfoque univariante: ANOVA con *k* grupos 380
Un enfoque multivariante: MANOVA con *k* grupos 383
Ejemplo 3: Diseño factorial del MANOVA con dos variables independientes 387
Primer paso: Objetivos del MANOVA 388
Segundo paso: Diseño de investigación del MANOVA 388
Tercer paso: Supuestos básicos del MANOVA 390
Cuarto paso: Estimación del modelo MANOVA y valoración del ajuste global 390
Quinto paso: Interpretación de los resultados 391
Una consideración general de los resultados 393
Resumen 396
Cuestiones de revisión 397
Bibliografía 397
Lectura complementaria: *Imagen y atracción de centros comerciales suburbanos*, por Francisco José Más Ruiz 399

Capítulo 7. Análisis conjunto 407

¿Qué es el análisis conjunto? 408
Un ejemplo hipotético de análisis conjunto 409
Un ejemplo empírico 410
La utilización del análisis conjunto en la práctica total 414
Comparación del análisis conjunto con otros métodos multivariantes 415

Técnicas de composición frente a las técnicas de descomposición ... 415
Especificación del valor teórico en el análisis conjunto ... 415
Modelos separados para cada individuo ... 415
Tipos de relaciones ... 416
Diseño de un experimento de análisis conjunto ... 416
Primer paso: Objetivos del análisis conjunto ... 417
Definición del valor total del objeto ... 419
Especificación de los factores determinantes ... 419
Segundo paso: Diseño del análisis conjunto ... 420
Selección de una metodología para el análisis conjunto ... 420
Diseño de los estímulos ... 421
Especificación de la forma básica del modelo ... 424
Recogida de datos ... 427
Tercer paso: Supuestos básicos del análisis conjunto ... 433
Cuarto paso: Estimación del modelo conjunto y valoración del ajuste global ... 433
Selección de una técnica de etimación ... 433
Evaluación de la bondad del ajuste del modelo ... 434
Quinto paso: Interpretación de los resultados ... 435
Análisis agregado frente a análisis desagradado ... 435
Evaluación de la importancia relativa de los relativos ... 436
Sexto paso: Validación de los resultados del análisis conjunto ... 436
Aplicación de los resultados del análisis conjunto ... 436
Segmentación ... 437
Análisis de rentabilidad ... 437
Simuladores conjuntos ... 437
Metodologías alternativas en el análisis conjunto ... 438
Análisis conjunto adaptativo: Análisis adaptativo con un gran número de factores ... 438
Resumen de las tres medologías completas ... 443
Una ilustración del análisis conjunto ... 444
Primer paso: Objetivos del análisis conjunto ... 444
Segundo paso: Diseño del análisis conjunto ... 444
Tercer paso: Supuestos básicos del análisis conjunto ... 446
Cuarto paso: Estimación del modelo conjunto y valoración del ajuste conjunto ... 446
Quinto paso: Interpretación de los resultados ... 448
Sexto paso: Validación de los resultados ... 449
Una aplicación práctica: Utilización de un simulador de elección ... 449
Resumen ... 450
Cuestiones de revisión ... 451
Bibliografía ... 451
Lectura complementaria: *Las preferencias del consumidor: estudio de su composición a través del análisis conjunto*, por Salvador Ruiz de Maya y José Luis Munuera Alemán ... 455

Capítulo 8. Análisis de correlación canónica ... 469

¿Qué es la correlación canónica? ... 470
Ejemplo hipotético de correlación canónica ... 470
El análisis de las relaciones mediante correlación canónica ... 471
Primer paso: Objetivos del análisis de correlación canónica ... 472
Segundo paso: Diseño de un análisis de correlación canónica ... 473
Tercer paso: Supuestos básicos de la correlación canónica ... 473
Cuarto paso: Obtención de las funciones canónicas y valoración del ajuste global ... 474
La derivación de funciones canónicas ... 474
¿Qué funciones canónicas se deben interpretar? ... 475
Quinto paso: Interpretación del valor teórico canónico ... 478

Ponderaciones canónicas 478
Cargas canónicas 478
Cargas cruzadas canónicas 479
El enfoque de interpretación a emplear 479
Sexto paso: Validación y diagnosis 479
Un ejemplo ilustrativo 480
Primer paso: Objetivos del análisis de correlación canónica 480
Segundo y tercer pasos: Diseño de un análisis de correlación canónica y contrastación de los supuestos . 481
Cuarto paso: Obtención de las funciones canónicas y valoración del ajuste global 481
Quinto paso: Interpretación de los valores teóricos canónicos 483
Sexto paso: Validación y diagnosis 486
Una consideración general para los resultados 486
Resumen 487
Cuestiones de revisión 487
Bibliografía 487

Sección 3. Técnicas de interdependencia 489

Capítulo 9. Análisis cluster 491

¿Qué es el análisis *cluster*? 492
¿Cómo funciona el análisis *cluster*? 493
Medición de la similitud 494
Formación de conglomerados 495
Determinación del número de conglomerados en la solución final 497
Proceso de decisión con el análisis *cluster* 498
Primer paso: Objetivos del análisis *cluster* 498
Selección de variables del análisis *cluster* 500
Segundo paso: Diseño de investigación mediante análisis *cluster* 500
Detección de atípicos 501
Medidas de similitud 502
Tipificación de los datos 506
Tercer paso: Supuestos del análisis *cluster* 507
Representatividad de la muestra. 508
Impacto de la multicolinealidad 508
Cuarto paso: Obtención de conglomerados y valoración del ajuste conjunto 508
Algoritmo para la obtención de conglomerados 509
¿Cuántos grupos deben formarse? 515
¿Debe volverse a especificar el análisis *cluster*? 516
Quinto paso: Interpretación de los conglomerados 516
Sexto paso: Validación y perfil de los grupos 517
Validación de la solución *cluster* 517
Perfiles de la solución *cluster* 518
Resumen del proceso de decisión 518
Un ejemplo ilustrativo 518
Primer paso: Objetivos del análisis *cluster* 518
Segundo paso: Diseño de investigación del análisis *cluster* 519
Tercer paso: Supuestos del análisis *cluster* 519
Cuarto paso: Obtención de grupos y valoración del ajuste conjunto 519
Quinto paso: Interpretación de los conglomerados 525
Sexto paso: Validación y perfil de los grupos 528
Una visión práctica 531
Resumen 531
Cuestiones de revisión 532

Bibliografía ... 532
Lecturas complementarias: *Aplicación del análisis cluster*, por Azucena Penelas y *El análisis cluster en investigación de marketing: Metodología y crítica*, por Mónica Gómez Suárez ... 533

Capítulo 10. Análisis multidimensional ... 547

¿Qué es el análisis multidimensional? ... 548
Una visión simplificada del análisis multidimensional ... 549
Comparación del análisis multdimensional con otras técnicas de interdependencia ... 551
El individuo como unidad de análisis ... 551
Ausencia de valor teórico ... 552
Un marco de decisión para la elaboración de mapas perceptuales ... 552
Primer paso: Objetivos del análisis multidimensional ... 552
Decisiones clave en la fijación de objetivos ... 552
Segundo paso: Diseño de investigación mediante el análisis multidimensional ... 555
Selección de un enfoque de descomposición (libre de atributos) o de composición (basada en atributos) ... 555
Objetos: Su número y selección ... 557
Métodos métricos frente a métodos no métricos ... 558
Recogida de datos sobre similitudes o preferencias ... 558
Tercer paso: Supuestos del análisis multidimensional ... 561
Cuarto paso: Obtención de soluciones de análisis multidimensional y valoración del ajuste conjunto ... 561
Determinación de la posición de un objeto en un mapa perceptual ... 561
Seleccionando la dimensionalidad del mapa perceptual ... 563
Incorporación de las preferencias en el análisis multidimensional ... 565
Quinto paso: Interpretación de los resultados del análisis multidimensional ... 569
Identificación de las dimensiones ... 569
Sexto paso: Validación de los resultados del análisis multidimensional ... 570
Análisis de correspondencias ... 571
Un ejemplo sencillo de CA ... 571
Primer paso: Objetivos del análisis de correspondencias ... 575
Segundo paso: Diseño de la investigación mediante análisis de correspondencias ... 576
Tercer paso: Supuestos del análisis de correspondencia ... 576
Cuarto paso: Obtención de resultados con análisis de correspondencias y valoración del ajuste conjunto ... 576
Quinto paso: Interpretación de los resultados ... 577
Sexto paso: Validación de los resultados ... 577
Visión de conjunto del análisis de correspondencias ... 577
Ilustración del análisis multidimensional y del análisis de correspondencias ... 578
Primer paso: Objetivos del mapa perceptual ... 578
Segundo paso: Diseño de la investigación mediante el estudio del mapa perceptual ... 578
Tercer paso: Supuestos básicos en la elaboración de un mapa perceptual ... 578
Análisis multidimensional: Pasos cuarto y quinto ... 580
Revisión de los resultados de descomposición ... 587
Análisis de correspondencias: Pasos cuarto y quinto ... 588
Sexto paso: Validación de los resultados ... 591
Un resumen para la dirección de los resultados del análisis multidimensional ... 592
Resumen ... 593
Cuestiones de revisión ... 593
Bibliografía ... 594
Lectura complementaria: *Aplicación del análisis multidimensional y del análisis factorial de correspondencias*, por Enrique Bigné Alcañiz y Natalia Vila López ... 595

Sección 4. Técnicas avanzadas emergentes ... 609

Capítulo 11. Modelos de ecuaciones estructurales ... 611

¿Qué es el modelo de ecuaciones estructurales? ... 612

Tratamiento de relaciones múltiples de dependencia cruzada 613
Incorporación de variables que no se miden directamente 613

Un ejemplo sencillo de SEM 614

La cuestión de investigación 615
Identificación del modelo de ecuación estructural para el análisis de secuencias 615
Una aplicación del análisis de secuencias 616
Resumen 617

La función de la teoría en los modelos de ecuaciones estructurales 617
Desarrollo de una estrategia de modelización 618

Estrategia de modelización confirmatoria 618
Estrategia de modelos rivales 619
Estrategia de desarrollo del modelo 619

Pasos en la modelización de ecuaciones estructurales 620

Primer paso: Desarrollo de un modelo basado en la teoría 620
Segundo paso: Construcción de un diagrama de secuencias de relaciones causales 622
Tercer paso: Conversión de un diagrama de secuencias en un conjunto de ecuaciones estructurales y especificación del modelo de medida 624
Cuarto paso: Selección del tipo de matriz de entrada y estimación del modelo propuesto 628
Quinto paso: Valoración de la identificación del modelo estructural 634
Sexto paso: Evaluación de los criterios de calidad de ajuste 637
Séptimo paso: Interpretación y modificación del modelo 640
Una recapitulación del proceso de siete pasos 642

Dos ilustraciones de modelización de ecuaciones estructurales 642
Análisis factorial confirmatorio 643

Primer paso: Desarrollo de un modelo basado en la teoría 643
Segundo paso: Construcción de un diagrama de relaciones causales 643
Tercer paso: Conversión de un diagrama de relaciones en un conjunto de modelos de ecuaciones estructurales y de medida 644
Cuarto paso: Elección del tipo de matriz de entrada y estimación del modelo propuesto 645
Quinto paso: Valoración de la identificación del modelo estructural 645
Sexto paso: Evaluación de criterios de calidad del ajuste 646
Séptimo paso: Interpretación y modificación del modelo 651
Modelos de análisis factorial de orden superior 652
Resumen 653

Estimación de un modelo de relaciones con SEM 653

Primer paso: Desarrollo de un modelo basado en la teoría 654
Segundo paso: Construcción de un diagrama de relaciones causales 655
Tercer paso: Conversión del diagrama de relaciones en un conjunto de ecuaciones estructurales y especificación del modelo de medida 655
Cuarto paso: Elección del tipo de matriz de entrada y estimación del modelo propuesto 656
Quinto paso: Evaluación de la identificación del modelo estructural 657
Sexto paso: Evaluación de los criterios de calidad del ajuste 657
Séptimo paso: Interpretación y modificación del modelo 666
Repaso del proceso de siete pasos 669

Resumen 669
Cuestiones de revisión 669

Apéndice 11A. Una representación matemática en notación LISREL 671

Notación LISREL 671
Desde un diagrama de relaciones a la notación LISREL 673

Construcción de las ecuaciones estructurales desde el diagrama de relaciones 674

Resumen 677

Apéndice 11B. Medidas conjuntas de calidad del ajuste para modelos de ecuaciones estructurales 679
Medidas del ajuste absoluto 680
Estadístico-ratio de verosimilitud chi-cuadrado 680
Parámetros de no centralidad y no centralidad a escala 681
Índice de calidad del ajuste 681
Residuo cuadrático medio 681
Error de aproximación cuadrático medio 682
Índice de validación cruzada esperada 682
Índice de validación cruzada 682
Medidas del ajuste incremental 682
Índice de calidad del ajuste ajustada 683
Índice Tucker-Lewis 683
Índice del ajuste normal 683
Otras medidas de ajuste incremental 683
Medidas de ajuste de parsimonia 683
Índice de ajuste normado de parsimonia 684
Índices de calidad de ajuste de parsimonia 684
Chi-cuadrado normada 684
Criterio de información de Akaike 685
Una revisión de las medidas de calidad del ajuste estructural 685
Resumen 685
Bibliografía 687
Lectura complementaria: *La confianza del consumidor en el minorista: Desarrollo y contraste de un modelo conceptual*, por Manuel Sánchez Pérez, M.ª Ángeles Iniesta y Elvira Sáez González 691

Capítulo 12. Nuevas técnicas del análisis multivariante 707
Introducción 708
La avalancha de información 708
Análisis sin inferencia estadística 709
Temas abarcados en este capítulo 709
Almacenamiento y extracción de datos 710
¿En qué consisten la extracción y el almacenamiento de datos? 711
Conceptos fundamentales del almacenamiento de datos 712
Supuestos fundamentales en la extracción de datos 714
Redes neuronales 721
Conceptos básicos de las redes neuronales 722
Estimación de un modelo de red neuronal 724
Utilización de una red neuronal para la clasificación 727
Resumen 728
Remuestreo 728
Un breve repaso de la inferencia paramétrica 729
Conceptos básicos de la reespecificación 729
Un ejemplo de reespecificación y regresión múltiple 730
Resumen 732
Resumen 733
Cuestiones de revisión 733
Bibliografía 734

Apéndice A. Aplicaciones del análisis multivariante 735

Glosario 767

Prólogo

¿Quién hubiera pensado que cuando se publicó la primera edición del Análisis Multivariante hace casi 20 años que el uso de las estadísticas multivariantes estaría tan omnipresente como ocurre hoy en día? Durante este tiempo hemos observado un cambio drástico en el entorno de investigación al que se enfrentan tanto los investigadores académicos como los empresariales. En primer lugar, la revolución del computador personal ha proporcionado un poder y que era inimaginable hace tan solo unos pocos años. En este tiempo hemos progresado desde las tarjetas perforadas al reconocimiento de la voz. También hemos observado avances importantes, tanto en la disponibilidad de programas estadísticos como en su facilidad de uso, que abarca desde los paquetes informáticos completamente integrados tales como el SPSS y el SAS hasta los programas especializados para técnicas tales como redes neuronales y análisis conjunto. Hoy en día, el investigador puede encontrar casi cualquier técnica concebible en un formato de computador personal y a un precio razonable.

Desde la perspectiva estadística, hemos observado un desarrollo continuo de las nuevas técnicas, tales como el análisis conjunto, los modelos de ecuaciones estructurales y las redes neuronales. Todos estos avances tecnológicos, sin embargo, se han visto igualados por una necesidad creciente de una capacidad más analítica. La explosión de datos de los últimos años no sólo ha hipotecado nuestros recursos para manejar físicamente y analizar toda la información disponible, sino que también ha obligado a una nueva valoración de nuestra aproximación al análisis de datos. Finalmente, la compleja naturaleza de los temas abordados y el creciente papel de la teoría en el diseño de la investigación se han unido para exigir técnicas más rigurosas y sofisticadas que permitan llevar a cabo los análisis confirmatorios necesarios.

Todos estos acontecimientos han contribuido a la aceptación de las últimas cuatro ediciones de este libro y han creado una demanda para esta quinta edición. Con esta revisión hemos intentado abordar las exigencias de académicos e investigadores aplicados, con una presentación que se basa fundamentalmente en las técnicas estadísticas. Sin embargo, también se centra en el diseño, la estimación y la interpretación. Nos esforzamos continuamente por reducir nuestra dependencia sobre la notación estadística y la terminología y en su lugar identificamos los conceptos fundamentales que tienen efecto sobre nuestro uso de estas técnicas y los expresamos en términos sencillos: una introducción de orientación básicamente práctica del análisis multivariante para personas que no son expertas en estadística. Continuamos con nuestro cometido de proporcionar una comprensión completa de los principios estadísticos y de gestión que subyacen en el análisis multivariante para desarrollar con comodidad no sólo las cuestiones estadísticas sino también las prácticas.

Lo que es nuevo

El cambio más obvio en la quinta edición es su reorganización en cuatro secciones. Esta organización sigue un esquema en paralelo más estrecho con el proceso de investigación, particularmente en el enfoque sobre la preparación de datos y el desarrollo de escalas antes de emplear las técnicas de dependencia o interdependencia. La Sección 1, «Preparación para un análisis multivariante» tiene como núcleo central la preparación de datos, tales como el análisis de datos ausentes y la valoración de supuestos estadísticos junto a la reducción de datos, con especial atención al desarrollo de las escalas aditivas. La Sección 2, «Técnicas de dependencia», se dedica a cinco técnicas de dependencia multivariante: la regresión múltiple, el análisis discriminante, la regresión logística, el análisis multivariante de la varianza, el análisis conjunto y la correlación canónica. La Sección 3, «Técnicas de interdependencia», proporciona una cobertura del análisis cluster y las escalas multidimensionales y sus aplicaciones en el uso de estructuras entre observaciones. La última sección, la Sección 4, «Técnicas avanzadas y emergentes» introduce los modelos de ecuaciones estructurales junto con algunas áreas nuevas del análisis multivariante, que incluyen la búsqueda y almacenaje de datos, las redes neuronales y la re-utilización de muestras.

Se han realizado dos incorporaciones a la explicación de cada técnica multivariante. En primer lugar, se proporciona un ejemplo sencillo al comienzo de cada capítulo para ilustrar los principios básicos y los objetivos y cómo funcionaría un método en particular en una situación real. En segundo lugar, a continuación de cada uno de los ejemplos que utilizan los datos HATCO, se ofrece una valoración general que proporciona una perspectiva sobre la interpretación de los resultados y cómo podrían emplearse para abordar la cuestión objeto de la investigación. Ambas incorporaciones proporcionan un contexto más ilustrativo para cada técnica.

Lo que ha sido desarrollado y actualizado

Cada capítulo ha sido revisado para incorporar los avances en la tecnología y varios capítulos han experimentado un cambio más extenso. Con la inclusión del Capítulo 3, «Análisis factorial» en la Sección 1, se ha hecho especial hincapié en el desarrollo de escalas aditivas y la utilización de las escalas o las puntuaciones factoriales en otras aplicaciones. El Capítulo 5, «Análisis discriminante múltiple y regresión logística» proporciona ahora una cobertura completa del análisis de variables dependientes categóricas con la inclusión tanto del análisis discriminante como de la regresión logística, y contiene un tratamiento desarrollado de otros modelos categóricos. El Capítulo 7, en la parte «Análisis conjunto» contiene una revisión de las cuestiones de diseño de investigación, que se centra en el desarrollo de los estímulos de una manera concisa y clara. Finalmente, el Capítulo 11, «Modelos de ecuaciones estructurales», ha sido actualizado para reflejar los múltiples cambios que ha sufrido este área durante los últimos años. Tres temas abordados son modelos factoriales de segundo orden; técnicas de estimaciones alternativas, tales como la simulación y «bootstrapping»; y algunos de los problemas operativos más comunes, tales como los datos ausentes o la obtención de una matriz definitiva no positiva durante una estimación de modelo. Cada uno de estos cambios, junto con otros que no han sido mencionados, ayudan a la adquisición de un conocimiento profundo de las cuestiones estadísticas y aplicadas que subyacen estas técnicas.

Con vistas al futuro

Hemos realizado otras dos incorporaciones que consideramos innovadoras y que aumentarán sustancialmente la comprensión del lector acerca del análisis multivariante. La primera es un capítulo nuevo —«Técnicas emergentes en el análisis multivariante». Conforme evoluciona el entorno investigador, los investigadores tienen que adaptarse a los cambios. El investigador de hoy día se

enfrenta a cantidades de información crecientes y la necesidad de un método de descubrimiento objetivo además de explicativo. Introducimos las temas del almacenaje y búsqueda de datos para exponer al investigador sus objetivos básicos y los principios implicados. Con la utilización de las perspectivas obtenidas en el debate de otras técnicas contrastamos la naturaleza exploratoria de la búsqueda de datos. También abordamos el concepto de redes neuronales, una de las técnicas estrechamente asociada con la búsqueda de datos y utilizada en muchas aplicaciones hoy en día. Otro tema es la reutilización de muestras, también conocido como «bootstrapping» y el efecto «jackknife». Esta aproximación ha estado disponible muchos años; pero no ha sido utilizada de forma general, y está obteniendo una aceptación muy amplia como alternativa a la estimación paramétrica. Los debates ofrecidos proporcionan una breve panorámica de los temas, así como ejemplos empíricos simples. Así demostramos la capacidad de las redes neuronales para llevar a cabo el análisis discriminante y los resultados de la utilización sobre muestras cuando se aplican a la regresión múltiple.

El último avance es la creación de una página Web http://www.prenhall.com/hair, dedicada al análisis multivariante y que la hemos llamado «Great Ideas in Teaching Multivariate Statistics». Esta página Web actúa como un centro de recursos para todos los que se interesen por el análisis multivariante, proporcionando conexiones para cada técnica y un foro para la identificación de nuevos temas o métodos estadísticos. De esta forma podemos ofrecer una contestación más personalizada a los investigadores en tiempo real, en vez de mediante una nueva edición del libro. También esperamos que esta Web sea un centro de entrada y salida de materiales sobre la enseñanza de estadística multivariante, proporcionando ejercicios, series de datos e ideas sobre proyectos.

Agradecimientos

Numerosas personas nos han ayudado en la realización de la quinta edición de este texto. Barbara Ross, doctora por la Universidad de Louisiana (Louisiana State University), nos ha provisto de imponderable ayuda en todas las fases de la revisión. Stern Neill, también doctor por la Universidad de Louisiana, proporcionó asistencia en el análisis de los datos y la adición de las nuevas técnicas estadísticas. Estamos en deuda con los siguientes supervisores por su inestimable ayuda en las adiciones a la quinta edición:

David Booth, Kent State University
Robert Bush, Memphis State University
Rabikar Chatterjee, Univerisity of Michigan
Kerri Curtis, Golden Gate University
Muzaffar Shaikh, Florida Institute of Technology

También queremos expresar nuestro agradecimiento a personas que colaboraron en anteriores ediciones. Fundamentalmente, Bruce Alford, University of Evansville; David Andrus, Kansas State University; Alvin C. Burns, Louisiana State University; Alan J. Bush, University of Memphis; Robert Bush, University of Memphis; Chaim Ehrman, University of Ilinois at Chicago; Joel Evans, Hofstra University; Thomas L. Gillpatrick, Portland State University; Dipak Jain, Northwestern University; John Lastovicka, University of Kansas; Margaret Liebman, La Salle University; Richard Netemeyer, Louisiana State University; Scott Roack, Northeeast Louisiana University; Walter A. Smith, Tulsa University; Ronald D. Taylor, Mississippi State University; y Jerry L.Wall, Northeast Louisiana University.

J. F. H.
R. E. A.
R. L. T
W. C. B.

Prólogo a la edición en castellano

La primera edición de este libro realizado por los autores Hair, Anderson y Tatham apareció en 1984. Desde entonces, hasta la 5.ª edición cuyo texto prologamos, se han producido profundas transformaciones en la teoría y práctica de la investigación de mercados que han sido recogidas fielmente por los autores.

La investigación de mercados inicia su expansión en la década de los años setenta con la incorporación de un conjunto de técnicas de análisis de datos procedente de la estadística. Hasta entonces, en España las fuentes doctrinales se basaban en la obra pionera de Tagliacarne (Ariel, 1962), estructurada en torno a la realización de encuestas y el muestreo. Posteriormente, en la década de los setenta se introduce en el mundo académico y empresarial un conjunto de técnicas estadisticas de análisis de datos en torno a los Cursos de Investigación Comercial de AEDEMO y de los Cuadernos Universitarios de Planificación Empresarial y Marketing (CUPEMA) que dirigió el Prof. Antonio Pulido.

Las décadas de los ochenta y noventa, con la difusión de los computadores personales y de programas informáticos más amigables para los usuarios, muestran la eclosión de la aplicación de las técnicas multivariantes a la investigación de mercados en las universidades y en las empresas, de las que los trabajos españoles que acompañan al texto original son un buen exponente.

Esta difusión y una mayor facilidad de utilización, que ofrece grandes ventajas, no están exentas de inconvenientes. Así, se puede mencionar la utilización excesiva de un conjunto de técnicas inadecuadas para el problema comercial objeto de estudio. La aplicación indiscriminada de las técnicas estadísticas a cualquier base de datos y la explicación posterior de los resultados obtenidos construyendo supuestas "teorías" explicativas ad-hoc, son uno de los principales errores. Todo investigador, empresarial o universitario, debe partir de una adecuada formulación del problema objeto de estudio, debe formular una teoría explicativa o, al menos, un conjunto de hipótesis y, finalmente, debe contrastar estas hipótesis con la evidencia empírica. La espectacularidad de las técnicas empleadas no puede nunca obviar la lógica de la investigación científica.

En la actividad universitaria constituye un punto de especial interés la delimitación de los campos de estudio de las diferentes disciplinas. En este sentido, es importante delimitar en la Investigación de Mercados el papel a desempeñar por la estadística como proveedora de las técnicas de análisis de la información comercial. En esta debatida cuestión, corresponde a la Investigación de Mercados la definición del problema objeto de estudio, la formulación de la hipótesis explicativas, la obtención de la evidencia empírica y la interpretación de los resultados, mientras que la Esta-

dística deberá suministrar el acerbo de conocimientos teóricos, procedimientos y algoritmos, que hagan posible las investigaciones planteadas.

El libro que hoy prologamos responde al enfoque antes expuesto. El tratamiento de las técnicas se hace desde una perspectiva del usuario, es decir, se explica cada método de acuerdo con los objetivos que podrían tenerse en cuenta al desarrollar una investigación, intentando no abrumar al lector con complejas notaciones estadísticas. Con este enfoque, los autores hacen énfasis en el desarrollo conceptual de la técnica y en la interpretación de los resultados, más que en el aparato estadístico que justifica cada método. No obstante, esto no quiere decir que las explicaciones y aplicaciones no sean rigurosas desde el punto de vista teórico, ya que siempre se trata de poner de manifiesto la importancia de determinados aspectos relevantes como el análisis previo de los datos, el tamaño muestral mínimo requerido para la aplicación de una técnica concreta, el cumplimiento de unos supuestos estadísticos básicos, la necesidad de que no sólo se analice el nivel de significación sino también la potencia estadística, el análisis de fiabilidad y validez de los resultados, etc.

La estructura del libro responde a un orden lógico que se basa en el diseño de modelos multivariantes para desarrollar las diferentes etapas a seguir en la aplicación de cada técnica. Así pues, cada capítulo dedicado a una técnica determinada sigue dichas etapas, de tal forma que el lector aprende el uso de los distintos métodos partiendo siempre de los objetivos de la investigación y finalizando con la significación práctica de los resultados. De esta forma se pone de manifiesto la importancia que tiene el diseño de la investigación como punto inicial de cualquier análisis estadístico. Por ello resulta una guía muy adecuada para los investigadores noveles que siempre tienen la tentación de justificar el uso de cualquier técnica, sin detenerse previamente a pensar si están utilizando las variables adecuadas o si se incumplen determinados supuestos que podrían comprometer gravemente los resultados de la investigación.

La labor de revisión de la traducción es una tarea que ha sido cuidadosamente realizada por Mónica Gómez Suárez, Profesora Titular de Comercialización e Investigación de Mercados en la Universidad Autónoma de Madrid, con una amplia experiencia en la docencia y la investigación. La edición anterior de este libro venía acompañada de un conjunto de lecturas referentes al mercado norteamericano, que han sido eliminadas en esta quinta edición en inglés. Como quiera que estas lecturas eran un buen complemento a los fundamentos teóricos del libro, se ha decidido, con gran acierto por parte de la editorial, la inclusión de ejemplos de publicaciones realizadas en España. La profesora Gómez Suárez ha seleccionado un conjunto de artículos representativos de cada una de las técnicas de investigación de mercados, que son buenos ejemplos de las investigaciones realizadas por los profesores de las universidades españolas y que muestran el alto nivel alcanzado por los investigadores españoles. Este esfuerzo constituye un importante valor añadido a la obra original.

Por todo ello pensamos que este libro va a constituir una referencia obligada entre los cada vez más numerosos y competentes estudiosos de la investigación de mercados.

En Madrid, a 30 de junio de 1999

Ignacio Cruz Roche
Catedrático de Comercialización e Investigación de Mercados
Universidad Autónoma de Madrid

Nota introductoria

Durante la década de los ochenta se fueron desarrollando los programas estadísticos que facilitaron el análisis de gran cantidad de datos cuyo origen estaba en encuestas o en bases de datos que provenían de fuentes secundarias de información. Los fundamentos teóricos o estadísticos de las técnicas multivariantes que permitían el análisis de estos datos habían sido desarrollados con anterioridad, pero sólo cuando los ordenadores tuvieron la capacidad de cálculo y memoria necesarios para llevar a cabo el análisis multivariante, se empezó a generalizar el uso de estas técnicas. La primera edición de este libro data de 1984, un año que nos parece ya bastante lejano si tenemos en cuenta la revolución que se ha experimentado en las técnicas de análisis de la información. En las ediciones posteriores, los autores han ido cambiando la estructura del libro, incluyendo nuevas técnicas y adaptando sus explicaciones a la práctica investigadora y empresarial.

Preparar y adaptar la quinta edición de "Análisis Multivariante" al mercado español ha supuesto una labor importante, puesto que no sólo se ha realizado la revisión técnica de la obra, sino también la selección de lecturas que complementan la comprensión de las técnicas mediante su aplicación a investigaciones realizadas en nuestro país. Hemos incluido artículos relacionados con el marketing publicados en revistas españolas y aplicaciones realizadas específicamente para este libro, de forma que participaran autores de diferentes universidades y que se tuviera en cuenta distintos ámbitos de aplicación de cada una de las técnicas. Por ello, en esta nota introductoria debemos destacar nuestro más sincero agradecimiento a estos autores que, no sólo han colaborado con sus investigaciones, sino que también han ayudado en la revisión de los capítulos correspondientes.

En el tercer capítulo, Análisis Factorial, hemos insertado la lectura de un artículo elaborado por el profesor Rodolfo Vázquez Casielles de la Universidad de Oviedo, que ilustra cómo se utiliza esta técnica cuando el objetivo es analizar la imagen proyectada por una serie de establecimientos minoristas y la percepción del servicio ofertado que tienen los consumidores.

El cuarto capítulo, Regresión Múltiple, se complementa con dos lecturas. La primera ha sido desarrollada por la Profesora M.ª Jesús Yagüe de la Universidad Autónoma de Madrid. En este caso, se utilizan datos que provienen de fuentes secundarias para mostrar la relación existente entre concentración y precios en el comercio minorista. Asimismo, para ilustrar las técnicas avanzadas de regresión hemos incluido el artículo de tres profesoras de la Universidad de Zaragoza, Carmen Berné, Marta Pedraja y Pilar Rivera. A diferencia del artículo anterior, se trata de ver el análisis de precios en el mercado minorista de alimentación desde la perspectiva del consumidor mediante regresión por los métodos de estimación ULS y GLS.

En el capítulo cinco, dedicado al Análisis Discriminante Múltiple y la Regresión Logística, hemos seleccionado dos artículos. El primero ha sido elaborado por los profesores Carlos Flavián, Eva Martínez y Yolanda Polo, de la Universidad de Zaragoza, que aplican un análisis discriminante

y un análisis logit para determinar qué factores son los más influyentes en el nivel de equipamiento doméstico a partir de una encuesta realizada a familias residentes en el ámbito metropólitano de Zaragoza. Con el objetivo de mostrar más específicamente cómo se utiliza el análisis logit multinomial, se incluye también el artículo de Sandra Cavero y Javier Cebollada, de la Universidad Pública de Navarra, cuya investigación tiene como objetivo analizar las estrategias de captación y retención de clientes por parte de las empresas. Los datos que sirven para realizar este análisis provienen de un panel de hogares.

El artículo que ilustra el capítulo seis, Análisis Multivariante de la Varianza, ha sido elaborado por Francisco Mas, de la Universidad de Alicante. El objetivo fundamental de su trabajo consiste en examinar las percepciones de los consumidores sobre los centros comerciales, así como la relación existente entre dichas percepciones y la atracción de los mismos. La información se obtiene mediante una encuesta y el análisis multivariante de la varianza se utiliza para probar si los niveles promedios de cada una de las dimensiones de imagen difieren entre los centros comerciales.

La lectura seleccionada para el capítulo siete, Análisis Conjunto, se basa en la investigación realizada por Salvador Ruiz de Maya y José Luis Munuera sobre las preferencias de consumidor. En el artículo se establecen previamente las etapas que se deben seguir al aplicar esta técnica y luego se muestra su utilización en un producto concreto, las marcas de colonia.

Para los capítulos nueve y diez hemos contado con la colaboración de tres profesores que han elaborado las aplicaciones de las técnicas específicamente para ser expuestas en este libro. Por ello, queremos agradecer el esfuerzo adicional que han tenido que desarrollar adaptando sus investigaciones a la exposición de dichas aplicaciones. En el capítulo dedicado a Análisis Cluster, Azucena Penelas, de la Universidad de Alcalá, ilustra la utilización de esta técnica para definir cómo las entidades bancarias españolas desarrollan sus estrategias de segmentación de mercados. En el capítulo diez, dedicado a Análisis Multidimensional y Factorial de Correspondencias, Enrique Bigné (Universidad Jaume I, Castellón) y Natalia Vila (Universidad de Valencia) muestran con fines exclusivamente pedagógicos para qué sirven del análisis multidimensional, el análisis factorial de correspondencias, el análisis de preferencias y el análisis de propiedades, para medir el posicionamiento a través de mapas perceptuales de cada uno de los cinco puntos de venta que compiten en dos localidades. En ambos capítulos, los datos se obtienen a partir de encuestas (la primera aplicación a directivos bancarios y la segunda, a amas de casa). También se incluye en el capítulo nueve, dedicado a análisis cluster, una revisión crítica y metodológica de la aplicación del análisis cluster en los años 90 en la disciplina de Marketing, mostrándose cómo se ha utilizado esta técnica y sus principales aplicaciones.

Por último, en el capítulo once, dedicado a modelos de ecuaciones estructurales, se complementa con el artículo de los profesores Manuel Sánchez, M.ª Angeles Iniesta y Elvira Sáez, de la Universidad de Almería, que plantean un modelo conceptual de la confianza en el contexto de las relaciones minorista-consumidor, a partir de un análisis factorial confirmatorio de orden superior.

En los capítulos ocho (Correlación Canónica) y doce (Técnicas emergentes) no hemos incluido ninguna lectura complementaria, puesto que la primera de las técnicas se ha utilizado muy poco en marketing tanto internacionalmente como en España y las últimas no cuentan todavía con desarrollos académicos que permitan ilustrar su aplicación. Respecto a este último capítulo, debemos destacar la colaboración de Jean Pierre Levy, de la Universidad de Quebec, que nos ha ayudado a hacer la revisión de este capítulo y nos ha proporcionado certeros consejos en cuanto a la traducción de algunos términos.

Antes de terminar, quiero agradecer el apoyo recibido de todos los profesores del Departamento de Financiación e Investigación Comercial de la Universidad Autónoma de Madrid y expresar mi satisfacción por haber podido contribuir a la difusión en lengua española de este libro.

En Madrid, a 30 de junio de 1999

Mónica Gómez Suárez
Profesora titular de Comercialización e Investigación de Mercados
Universidad Autónoma de Madrid

Profesores que han participado con lecturas complementarias

Rodolfo Vázquez Casielles	*Universidad de Oviedo*
M.ª Jesús Yagüe Guillén	*Universidad Autónoma de Madrid*
Carmen Berné Manero	*Universidad de Zaragoza*
Marta Pedraja Iglesias	*Universidad de Zaragoza*
Pilar Rivera Torres	*Universidad de Zaragoza*
Carlos Flavián Blanco	*Universidad de Zaragoza*
Eva Martínez Salinas	*Universidad de Zaragoza*
Yolanda Polo Redondo	*Universidad de Zaragoza*
Sandra Cavero Brújula	*Universidad Pública de Navarra*
Javier Cebollada Calvo	*Universidad Pública de Navarra*
Francisco José Más Ruiz	*Universidad de Alicante*
Salvador Ruiz de Maya	*Universidad de Murcia*
José Luis Munuera Alemán	*Universidad de Murcia*
Azucena Penelas	*Universidad de Alcalá de Henares*
Mónica Gómez Suárez	*Universidad Autónoma de Madrid*
Enrique Bigné Alcañiz	*Universitat Jaume I de Castellón*
Natalia Vila López	*Universitat de Valencia*
Manuel Sánchez Pérez	*Universidad de Almería*
M.ª Ángeles Iniesta	*Universidad de Almería*
Elvira Sáez González	*Universidad de Almería*

CAPÍTULO 1

Introducción

OBJETIVOS DE APRENDIZAJE

- Explicar qué es el análisis multivariante y cuándo es apropiada su aplicación.
- Definir y comentar las técnicas concretas incluidas en el análisis multivariante.
- Determinar qué técnica multivariante es la indicada para un problema de investigación específico.
- Explicar la naturaleza de las escalas de medida y su relación con las técnicas multivariantes.
- Describir los aspectos conceptuales y estadísticos propios del análisis multivariante.

El Capítulo 1 constituye una revisión simplificada de las técnicas del análisis multivariante. Se resalta la influencia creciente de los métodos del análisis multivariante, no sólo en los aspectos analíticos de la investigación, sino también en el diseño y forma en que se recogen datos para la adopción de decisiones y la resolución de problemas. Aunque las técnicas multivariantes comparten bastantes características con sus análogas uni y bivariantes, son evidentes las diferencias que existen al pasar al ámbito del análisis multivariante. Para ilustrar éstas, este capítulo presenta una clasificación de las técnicas multivariantes. Después, proporciona unas indicaciones generales para la aplicación de estas técnicas junto con una aproximación estructurada a su formulación, estimación y a la interpretación de resultados multivariantes. El capítulo concluye con un comentario sobre la base de datos utilizada en la mayor parte del libro para ilustrar las aplicaciones de las técnicas.

¿Qué es el análisis multivariante?

La tecnología informática disponible hoy en día, casi inimaginable hace sólo dos décadas, ha hecho posible avances extraordinarios en el análisis de datos psicológicos, sociológicos y de otro tipo de datos referidos al comportamiento humano. Este impacto es más evidente en la relativa facilidad con la que los ordenadores pueden analizar enormes cantidades de datos complejos. Casi cualquier problema se puede analizar fácilmente hoy en día por un número ilimitado de programas estadísticos, incluso en ordenadores personales. Además, los efectos del progreso tecnológico han extendido aún más la capacidad de manipular datos, liberando a los investigadores de las restricciones del pasado y permitiéndoles así abordar investigaciones más sustantivas y ensayar sus modelos teóricos. Las limitaciones metodológicas no son ya un asunto crítico para el teórico empeñado en la búsqueda de evidencia empírica. Gran parte de esta creciente comprensión y pericia en el análisis de datos ha venido a través del estudio de la estadística y de la inferencia estadística. Igualmente importante, sin embargo, ha sido el dilatado conocimiento y aplicación de un grupo de técnicas estadísticas conocidas como **análisis multivariante**.

Las técnicas del análisis multivariante están siendo ampliamente aplicadas a la industria, administración y centros de investigación de ámbito universitario. Por otra parte, pocos campos de investigación o estudio han fracasado en integrar las técnicas multivariantes en su «caja de herramientas» analítica. Para atender este creciente interés, se han publicado numerosos libros y artículos sobre los aspectos teóricos y matemáticos de estas herramientas, y han aparecido textos introductorios en casi todos los campos. Sin embargo, se han escrito pocos libros para el investigador que no es un especialista en matemáticas o estadística. Y menos aún que traten las *aplicaciones* de la estadística multivariante así como que suministren un tratamiento conceptual de los métodos estadísticos. Este libro se ha escrito para cubrir este vacío.

Los libros orientados a las aplicaciones son de un interés crucial para los estudiosos del comportamiento, los directivos de empresa o los funcionarios de la administración pública, sea cual sea su entorno, que tienen que desarrollar sus conocimientos del análisis multivariante para conseguir una mejor comprensión de los complejos fenómenos de sus ámbitos de trabajo. Cualquier investigador que examina sólo relaciones de dos variables y que evita el análisis multivariante está ignorando poderosas herramientas que podrían suministrar información potencialmente útil. Como afirma un investigador, «para los propósitos de... cualquier... disciplina aplicada, la mayor parte de nuestras herramientas son, o deberían ser, multivariantes. Uno llega a una conclusión que a menos.... que el problema sea tratado como un problema multivariante, está tratado superficialmente». Según los estadísticos Hardyck y Petrinovich:

> El análisis de los métodos multivariantes predominará en el futuro y dará por resultado cambios drásticos en el modo en que los investigadores piensan sobre los problemas y en cómo diseñan sus investigaciones. Esos métodos hacen posible plantear preguntas específicas y precisas de considerable complejidad en marcos idóneos, lo que posibilita llevar a cabo investigaciones teóricamente significativas y evaluar los efectos de las variaciones paramétricas ocurridas de forma natural en el contexto en que normalmente ocurren. De esta forma, se pueden preservar las correlaciones naturales entre las múltiples influencias sobre el comportamiento y se pueden estudiar estadísticamente los efectos aislados de esas influencias sin provocar el típico aislamiento de esos individuos o variables.

Por ejemplo, en la mayor parte de los mercados actuales, los directivos no pueden fiarse de las antiguas aproximaciones donde se consideraban consumidores homogéneos y caracterizados por un reducido número de variables demográficas. En su lugar, deben desarrollar estrategias para atraer a numerosos segmentos de clientes con diversas características demográficas y psicográficas en un mercado con múltiples restricciones (legales, económicas, competitivas, tecnológicas, etc.). Sólo a través de las técnicas de análisis multivariante se pueden examinar adecuadamente las relaciones múltiples de este tipo para llegar a una comprensión de la toma de decisiones más completa y realista.

En todo el texto usamos el término «investigador» cuando nos referimos a un analista, tanto en la comunidad empresarial como académica. No consideramos oportuno hacer una distinción entre estas dos áreas, puesto que la investigación en cualquiera de las dos tiene que depender de bases teóricas y cuantitativas. Aunque los objetivos de la investigación y el énfasis en la interpretación pueden variar, un investigador tiene que abordar todos los asuntos, tanto conceptuales como empíricos, que surgen en los debates de los métodos estadísticos.

El impacto de la revolución informática

Es poco menos que imposible discutir la aplicación de las técnicas multivariantes sin una mención al impacto de la informática. Como se ha expuesto antes, el amplio desarrollo de la aplicación de los computadores (primero el computador y más recientemente los computadores personales o los microcomputadores) para procesar grandes y complejas bases de datos, ha estimulado de manera impresionante el uso de los métodos de estadística multivariante. Toda la estadística teórica de las técnicas multivariantes actuales fue desarrollada mucho antes de la aparición de los computadores, pero sólo cuando estuvo disponible el poder de la informática para realizar cálculos cada vez más complejos llegó a conocerse la existencia de esas técnicas fuera del círculo de los estadísticos teóricos. Los continuos avances tecnológicos en informática, particularmente en los computadores personales, han puesto a disposición de cualquier investigador interesado el acceso a todos los recursos necesarios para resolver un problema multivariante de casi cualquier dimensión. De hecho, muchos investigadores se llaman a sí mismos *analistas de datos* en lugar de estadísticos o (en lenguaje llano) «aficionados a lo cuantitativo». Estos analistas de datos han contribuido sustancialmente al aumento del uso y aceptación de la estadística multivariante en los negocios y en la administración. En la comunidad académica, disciplinas de todos los campos del saber han adoptado las técnicas multivariantes, y los académicos deben estar cada vez más versados en las técnicas multivariantes apropiadas para sus investigaciones empíricas. Incluso para personas con sólida preparación cuantitativa, la disponibilidad de programas preparados para el análisis multivariante ha facilitado la compleja manipulación de matrices de datos que durante mucho tiempo ha retrasado el crecimiento de técnicas multivariantes.

Muchas de las principales universidades americanas están pidiendo ya a los estudiantes como requisito de entrada comprar sus propios computadores incluso antes de su matriculación. Los estudiantes y profesores analizan rutinariamente datos multivariantes para responder a cuestiones surgidas en disciplinas de la antropología y la zoología. Todos los programas completos de estadística diseñados para computadores (por ejemplo, SPSS, SAS y BMDP) ahora se encuentran también para computadores personales. Los programas especializados para todo tipo de análisis multivariante, incluyendo los de escala multidimensional, modelos de ecuaciones simultáneas/estructurales y análisis conjunto, en su día se ofrecieron sólo para computadores siendo hoy en día compatibles con los computadores personales. Se están desarrollando sistemas expertos dirigidos incluso a temas tales como la selección de una técnica estadística [4] o diseñar un plan de muestreo que asegure los objetivos prácticos y estadísticos deseados [3].

Los programas estadísticos ya no se desarrollan primero en sistemas de computadores centrales para después trasladarlos a computadores personales o microcomputadores, sino que se desarrollan directamente para computadores personales. Quizá la categoría de programas de estadística de mayor desarrollo sean los paquetes estadísticos diseñados específicamente para aprovecharse de la flexibilidad del computador personal. Las técnicas multivariantes están tan extendidas que *todas* las técnicas ilustradas en este texto pueden ser estimadas con paquetes estadísticos fácilmente accesibles tanto para un computador, un microcomputador o un computador personal. La lista completa de los programas principales disponibles de software para el análisis multivariante se puede encontrar en el Apéndice A. Es necesario prestar una atención especial a los programas para el computador personal.

Definición del análisis multivariante

El análisis multivariante no es fácil de definir. En un sentido amplio, se refiere a todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación. Cualquier análisis simultáneo de más de dos variables puede ser considerado aproximadamente como un análisis multivariante. En sentido estricto, muchas técnicas multivariantes son extensiones del análisis univariante (análisis de distribuciones de una sola variable) y del análisis bivariante (clasificaciones cruzadas, correlación, análisis de la varianza y regresiones simples utilizadas para analizar dos variables). Por ejemplo, una regresión simple (con una variable predictor) se extiende al caso multivariante para incluir varias variables predictor. De la misma forma, la variable dependiente que se encuentra en el análisis de la varianza se extiende para incluir múltiples variables dependientes en el análisis multivariante de la varianza. Como veremos más adelante, en muchas ocasiones las técnicas multivariantes son un medio de representar en un análisis simple aquello que requirió varios análisis utilizando técnicas univariantes. Otras técnicas multivariantes, sin embargo, están diseñadas exclusivamente para tratar con problemas multivariantes, tales como el análisis factorial que sirve para identificar la estructura subyacente de un conjunto de variables o el análisis discriminante que sirve para diferenciar entre grupos basados en un conjunto de variables.

Una de las razones de la dificultad de definir el análisis multivariante es que el término multivariante no se usa de la misma forma en la literatura. Para algunos investigadores, *multivariante* significa simplemente examinar relaciones entre más de dos variables. Otros usan el término sólo para problemas en los que se supone que todas las variables múltiples tienen una distribución normal multivariante. Sin embargo, para ser considerado verdaderamente multivariante, todas las variables deben ser aleatorias y estar interrelacionadas de tal forma que sus diferentes efectos no puedan ser interpretados separadamente con algún sentido. Algunos autores afirman que el propósito del análisis multivariante es medir, explicar y predecir el grado de relación de los *valores teóricos* (combinaciones ponderadas de variables). Por tanto, el carácter multivariante reside en los múltiples *valores teóricos* (combinaciones múltiples de variables) y no sólo en el número de variables u observaciones. A efectos de este libro, no insistiremos en una definición rígida del análisis multivariante. En lugar de esto, el **análisis multivariante** incluirá tanto técnicas multivariables como técnicas multivariantes, debido a que los autores creen que el conocimiento de las técnicas multivariables es un primer paso esencial en la comprensión del análisis multivariante.

Algunos conceptos básicos del análisis multivariante

Aunque el análisis multivariante tiene sus raíces en la estadística univariante y bivariante, la extensión al dominio multivariante introduce conceptos y cuestiones adicionales. Estos conceptos van desde la necesidad de un entendimiento conceptual del elemento básico del análisis multivariante —el valor teórico— a las cuestiones específicas acerca de los tipos de escalas de medida utilizadas y los resultados estadísticos de los test de significación y los intervalos de confianza. Cada concepto juega un papel importante en la correcta aplicación de cualquier técnica multivariante.

El valor teórico

Como ya se ha mencionado, el elemento esencial del análisis multivariante es el **valor teórico**, una combinación lineal de variables con ponderaciones determinadas empíricamente. El investigador especifica las variables, mientras que las ponderaciones son objeto específico de determinación por parte de la técnica multivariante. Un valor teórico de n variables ponderadas (X_1 a X_n) puede expresarse matemáticamente así:

$$\text{Valor teórico} = w_1X_1 + w_2X_2 + w_3X_3 + \ldots + w_nX_n$$

donde X_n es la variable observada y w_n es la ponderación determinada por la técnica multivariante.

El resultado es un valor único que representa una combinación de *todo el conjunto* de variables que mejor se adaptan al objeto del análisis multivariante específico. En regresiones múltiples, el valor teórico se determina de tal forma que guarde la mejor correlación con la variable que se está prediciendo. En el análisis discriminante, el valor teórico se forma de tal manera que produzca resultados para cada observación que diferencien de forma máxima entre grupos de observaciones. Y en el análisis factorial, los valores teóricos se forman para representar mejor las estructuras subyacentes o la dimensionalidad de las variables tal y como se representan en sus intercorrelaciones.

En cada caso, el valor teórico capta el carácter multivariante del análisis. Por tanto, en nuestras discusiones de cada técnica, el valor teórico es el punto central del análisis por varias razones. Debemos entender no sólo su impacto conjunto para lograr cumplir el objetivo de cada técnica, sino también la contribución de cada variable separada al efecto del valor teórico en su conjunto.

Escalas de medida

El análisis de los datos implica la separación, identificación y medida de la variación en un conjunto de variables, tanto entre ellas mismas como entre una variable dependiente y una o más variables independientes. El término clave aquí es *medida,* dado que el investigador no puede separar o identificar una variación a menos que pueda ser mesurable. La medida es importante para representar con precisión el concepto de nuestro interés y es crucial en la selección del método de análisis multivariante apropiado. En los siguientes párrafos vamos a discutir el concepto de medida en lo que se refiere al análisis de datos y particularmente a las diversas técnicas multivariantes.

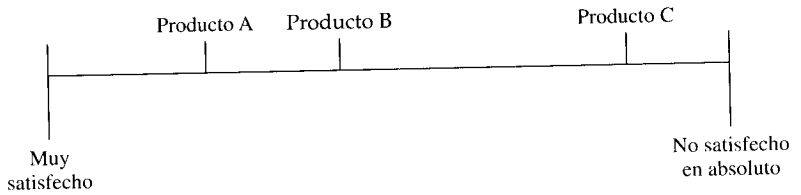
Existen dos tipos básicos de datos: **no métricos** (cualitativos) y **métricos** (cuantitativos). Los datos no métricos son atributos, características o propiedades categóricas que identifican o describen a un sujeto. Describen diferencias en tipo o clase indicando la presencia o ausencia de una característica o propiedad. Muchas propiedades son discretas porque tienen una característica peculiar que excluye todas las demás características. Por ejemplo, si uno es hombre, no puede ser mujer. No hay cantidad de «género», sólo la condición de ser hombre o mujer. Por el contrario, las medidas de datos métricos están constituidas de tal forma que los sujetos pueden ser identificados por diferencias entre grado o cantidad. Las variables medidas métricamente reflejan cantidades relativas o grado. Las medidas métricas son las más apropiadas para casos que involucran cantidad o magnitud, tales como el nivel de satisfacción o la demanda de trabajo.

Escalas de medida no métricas

Las medidas no métricas pueden tener escalas nominales u ordinales. La medida con una escala nominal asigna números que se usan para etiquetar o identificar sujetos u objetos. Las escalas nominales, también conocidas como escalas de categoría, proporcionan el número de ocurrencias en cada clase o categoría de la variable que se está estudiando. Por tanto, los números o símbolos asignados a los objetos no tienen más significado cuantitativo que indicar la presencia o ausencia del atributo o característica bajo investigación. Los ejemplos de datos con escala nominal incluyen el sexo, la religión o el partido político de una persona. Para trabajar con estos datos, el analista puede asignar números a cada categoría, por ejemplo, 2 para mujeres y 1 para hombres. Estos números sólo representan categorías o clases y no implican cantidades de un atributo o característica.

Las escalas ordinales representan un nivel superior de precisión de la medida. Las variables pueden ser ordenadas o clasificadas con escalas ordinales en relación a la cantidad del atributo poseído. Cada subclase puede ser comparada con otra en términos de una relación de «mayor que» o «menor que». Por ejemplo, los diferentes niveles de satisfacción del consumidor individual con diferentes productos nuevos puede ilustrarse en una escala ordinal. La siguiente escala muestra la idea

que tiene un encuestado acerca de tres productos. El encuestado está más satisfecho con A que con B y más satisfecho con B que con C.



Los números utilizados en escalas ordinales como éstas no son cuantitativos, dado que indican sólo posiciones relativas en series ordenadas. No hay medida de cuánta satisfacción recibe el consumidor en términos absolutos, el investigador ni conoce la diferencia exacta entre puntos de la escala de satisfacción. Muchas escalas de las ciencias del comportamiento caen dentro de esta categoría ordinal.

Escalas de medida métrica

Las escalas de intervalos y de razón (ambas métricas) proporcionan el nivel más alto de medida de precisión, permitiendo realizar casi todas las operaciones matemáticas. Estas dos escalas tienen unidades constantes de medida, de tal forma que las diferencias entre dos puntos adyacentes de cualquier parte de la escala son iguales. La única diferencia real entre las escalas de intervalo y las de razón es que las de intervalo tienen un punto cero arbitrario, mientras que las escalas de razón tienen un punto de cero absoluto. Las escalas de intervalo más familiares son las escalas de temperatura Celsius y Fahrenheit. Ambas tienen un punto de cero arbitrario, pero ese cero no indica una cantidad cero o ausencia de temperatura, dado que podemos registrar temperaturas por debajo del punto cero de esa escala. Por tanto, no es posible decir que un valor cualquiera situado en un intervalo de la escala es un múltiplo de cualquier otro punto de la escala. Por ejemplo, si un día se registran 80°F, no se puede decir que sea dos veces más caluroso que uno de 40°F porque sabemos que 80°F, en una escala diferente como Celsius, equivalen a 26,7°C. De la misma forma, 40°F en Celsius corresponden a 4,4°C. Aunque 80°F son, desde luego, dos veces 40°F, no se puede afirmar que el calor de 80°F sea dos veces el calor de 40°F porque usando diferentes escalas, el calor no es dos veces mayor; esto es, $4{,}4°F \times 2 \neq 26{,}7°C$.

Las escalas de razón representan la forma superior de medida de precisión, dado que poseen las ventajas de todas las escalas inferiores más un punto de cero absoluto. Con las medidas de escala de razón se permiten todas las operaciones matemáticas. El peso que tenemos en el baño u otras máquinas de peso comunes utilizan estas escalas, dado que tienen un punto de cero absoluto y que pueden ser expresados en términos de múltiplos cuando se relaciona un punto con otro de la escala; por ejemplo, 100 kilos es dos veces más pesado que 50 kilos.

Es importante entender los diferentes tipos de escalas de medida por dos razones. En primer lugar, el investigador debe identificar la escala de medida de cada variable empleada, de tal forma que no se estén utilizando datos no métricos como si fueran métricos. En segundo lugar, la escala de medida es crucial para determinar qué técnica multivariante es la más conveniente para los datos, consideración hecha tanto para las variables dependientes como las independientes. En la discusión de las técnicas y su clasificación, que haremos en posteriores secciones de este capítulo, las propiedades métricas o no métricas de las variables dependientes o independientes son los factores determinantes en la selección de la técnica apropiada.

Error de medida y medidas multivariantes

El uso de múltiples variables así como la dependencia de su combinación (el valor teórico) en las técnicas multivariantes también dirige su atención a un tema complementario, el error de medida.

El error de medida es el grado en que los valores observados no son representativos de los valores «verdaderos». El error de medida tiene múltiples fuentes, que van desde errores en la entrada de datos a la imprecisión en la medición (por ejemplo, imponiendo escalas de puntuación de siete puntos a la actitud medida cuando el investigador sabe que los encuestados sólo pueden responder con precisión a una puntuación de tres puntos) pasando por la incapacidad de los encuestados a proporcionar información precisa (por ejemplo, las respuestas a la renta de una economía familiar pueden ser razonablemente precisas pero rara vez lo son completamente). Por tanto, se debe asumir que todas las variable usadas en las técnicas multivariantes tienen algún grado de error de medida. El impacto del error de medida es añadir «ruido» a las variables medidas u observadas. Por tanto, el valor observado obtenido representa tanto el nivel «verdadero» como el «ruido». Cuando se calculan correlaciones o medias, normalmente el efecto «verdadero» está parcialmente camuflado por el error de medida, causando la debilidad de las correlaciones y la pérdida de precisión de las medias. El impacto específico del error de medida en las relaciones de dependencia se trata con más detalle en el Capítulo 11.

El objetivo del investigador de reducir el error de medida puede seguir varios caminos. Al valorar el grado de error de medida presente en cualquier medición, el analista debe enfrentarse tanto con la **validez** como con la **fiabilidad** de la medida. La validez es el grado en que la medida representa con precisión lo que se supone que representa. Por ejemplo, si queremos medir la renta discrecional, no preguntaremos por la renta total de las economías domésticas. Asegurar la validez empieza con un conocimiento profundo de lo que se va a medir y sólo entonces realizar la medida tan «correcta» y precisa como sea posible. Sin embargo, la precisión no asegura la validez. En nuestro ejemplo de la renta, el investigador podría definir muy precisamente el total de la renta familiar pero no tiene una medida válida de la renta discrecional porque no se ha planteado la pregunta «correcta».

Si la validez está asegurada, el investigador debe considerar la fiabilidad de las medidas. La fiabilidad es el grado en que la variable observada mide el valor «verdadero» y está «libre de error»; por tanto es lo opuesto al error de medida. Si la misma medida se realiza repetidas veces, por ejemplo, las medidas más fiables mostrarán una mayor consistencia que las medidas menos fiables. El investigador deberá valorar siempre las variables que están siendo usadas y si se pueden encontrar medidas alternativas válidas, elegir la variable con la mayor fiabilidad.

El investigador puede también optar por desarrollar **mediciones multivariantes**, también conocidas como **escalas sumadas**, donde diversas variables se unen en una **medida compuesta** para representar un concepto (por ejemplo, una escala de personalidad de entrada múltiple o puntuaciones sumadas de un producto). El objetivo es evitar usar sólo una única variable para representar un concepto, y en su lugar utilizar varias variables como **indicadores**, representando todos ellos diferentes facetas del concepto para obtener una perspectiva más completa. El uso de indicadores múltiples permite al investigador llegar a una especificación más precisa de las respuestas deseadas y no deja la fiabilidad plena a una única respuesta sino en la respuesta «media» o «típica» de un conjunto de respuestas relacionadas. Por ejemplo, al medir la satisfacción, uno podría preguntar una única cuestión, «¿cuál es su grado de satisfacción?», y basar el análisis en una única respuesta. O se podría desarrollar una escala aditiva que combinara varias respuestas de satisfacción, quizá en diferentes formatos de respuesta y en diferentes áreas de interés, que contemple la satisfacción total. La premisa básica es que las respuestas múltiples reflejan con mayor precisión la respuesta «verdadera» que la respuesta única. Para una introducción más detallada de los modelos de medidas múltiples y construcción de escalas, véase el Capítulo 3 (Análisis factorial) y el Capítulo 11 (Modelación mediante ecuaciones estructurales) o recursos adicionales [10]. Además, se han publicado en los últimos años compilaciones de escalas que proporcionan al investigador una escala «lista para ser empleada» con una fiabilidad demostrada [1, 5].

El impacto del error de medida y la escasa fiabilidad no pueden ser observadas directamente, dado que se encuentran en las variables observadas. El investigador debe, por tanto, trabajar siempre para aumentar la validez y la fiabilidad, lo que al final llevará a un retrato más «auténtico» de las variables de interés. Los malos resultados no siempre se deben al error de medida, pero la presencia del error de medida es garantía de distorsión en las relaciones observadas y hace menos poderosas las técnicas multivariantes. Reducir el error de medida, aunque implique esfuerzo, tiempo

y recursos adicionales, puede mejorar resultados débiles o marginales, así como fortalecer resultados probados.

Significación estadística frente a potencia estadística

Todas las técnicas multivariantes, excepto el análisis cluster y el análisis multidimensional, se basan en la inferencia estadística de los valores de una población o la relación entre variables de una muestra escogida aleatoriamente de esa población. Si estamos realizando un censo de toda la población, entonces la inferencia estadística no es necesaria, porque cualquier diferencia o relación, por pequeña que sea, es «verdad» y existe. Pero rara vez, casi nunca, se realiza un censo; por tanto, el investigador está obligado a deducir inferencias de una muestra.

Para interpretar las inferencias estadísticas, el investigador debe especificar los niveles aceptables de error estadístico. El modo de aproximación más común es determinar el nivel de **error de Tipo I**, también conocido como **alfa** (α). El error de Tipo I es la probabilidad de rechazar la hipótesis nula cuando es cierta, o expresado en términos más sencillos, la posibilidad de que la prueba muestre significación estadística cuando en realidad no está presente (el caso de un «positivo falso»). Especificando un nivel alfa, el investigador fija los márgenes admisibles de error especificando la probabilidad de concluir que la significación existe cuando en realidad no existe.

Al especificar el nivel de error de Tipo I, el investigador también determina un error asociado, denominado el **error de Tipo II** o **beta** (β). El error de Tipo II es la probabilidad de fallar en rechazar la hipótesis nula cuando es realmente falsa. Una probabilidad más interesante es $1 - \beta$, denominado la **potencia** del test de inferencia estadística. Potencia es la probabilidad de rechazar correctamente la hipótesis nula cuando debe ser rechazada. Por tanto, la potencia es la probabilidad de que la inferencia estadística se indique cuando esté presente. La relación de las diferentes probabilidades de error se muestra a continuación en el hipotético planteamiento de la evaluación de la diferencia entre dos medias:

		Realidad	
		H_0: Cierta	H_a: Falsa
Decisión estadística	H_0: Aceptar	$1 - \alpha$	β Error Tipo II
	H_a: No aceptar	α Error Tipo I	$1 - \beta$ Potencia

Aunque la especificación alfa establece el nivel de significación estadística aceptable, es el nivel de potencia el que dicta la probabilidad de «éxito» en la búsqueda de las diferencias si es que realmente existen. Entonces, ¿por qué no se plantean niveles aceptables tanto de alfa como de beta? Porque los errores de Tipo I y Tipo II están inversamente relacionados, y a medida que el error de Tipo I se hace más restrictivo (se acerca a cero), el error de Tipo II aumenta. Al disminuir el error de Tipo I también se reduce el poder de la prueba estadística. Por tanto, el analista tiene que conseguir un equilibrio entre el nivel de alfa y la potencia resultante.

La potencia no es sólo una función de alfa. Realmente está determinada por tres factores:

1. Efecto tamaño—La probabilidad de conseguir significación estadística se basa no sólo en consideraciones estadísticas sino también en la magnitud real del efecto que nos interesa (por ejemplo, una diferencia de medias entre dos grupos o la correlación entre variables) en la población, denominado **efecto tamaño**. Como cabría esperar, un efecto grande es más probable de encontrar que un efecto pequeño y por tanto, afecta a la potencia de la prueba estadística. Para evaluar la potencia de cualquier prueba estadística, el investigador debe entender primero el efecto examinado. Los efectos de tamaño se miden en términos estandarizados para facilitar la comparación. Las diferencias respecto de

la media se determinan en términos de desviaciones estándar, así que un efecto tamaño de 0,5 indica que la diferencia respecto de la media es la mitad de la desviación estándar. Para las correlaciones, el efecto tamaño se basa en la correlación efectiva entre las variables.

2. Alfa (α)—Como ya se ha discutido, a medida que alfa se vuelve más restrictivo, la potencia decrece. Esto significa que como el analista reduce la oportunidad de encontrar un efecto incorrecto significativo, la probabilidad de encontrar correctamente un efecto también disminuye. Las directrices convencionales sugieren niveles alfa de 0,05 o 0,01. Pero el investigador debe considerar el impacto de esta decisión sobre la potencia antes de seleccionar el nivel alfa. En discusiones posteriores se ilustrarán las relaciones entre las dos probabilidades.
3. El tamaño de la muestra—Para cualquier nivel de alfa dado, el aumento de la muestra siempre produce una mayor potencia del test estadístico. Pero aumentar el tamaño de la muestra también puede producir «demasiada» potencia. Por este hecho, entendemos que al aumentar el tamaño de la muestra, se observará que efectos cada vez más y más pequeños serán significativos, hasta que para muestras muy grandes casi cualquier efecto es significativo. El investigador debe tener siempre presente que el tamaño de la muestra puede afectar a la prueba estadística tanto por hacerlo insensible (para muestras muy pequeñas) o demasiado sensible (para muestras muy grandes).

Las relaciones entre alfa, tamaño de la muestra, efecto tamaño y potencia son bastante complicadas, pero se pueden encontrar ciertos puntos de partida. Cohen [6] ha examinado la potencia para la mayor parte de las pruebas de inferencia estadística y ha proporcionado pautas para los niveles aceptables de potencia, sugiriendo que los estudios deben diseñarse para conseguir niveles de alfa de al menos 0,05 con niveles de potencia del 80 por ciento. Para conseguir dichos niveles, deben considerarse simultáneamente los tres factores. Estas interrelaciones se pueden ilustrar mediante dos ejemplos sencillos. El primero implica la comprobación de la diferencia entre las puntuaciones medias de dos grupos. Suponiendo que el efecto tamaño sea entre pequeño (0,02) y moderado (0,5), el investigador debe determinar el nivel alfa y el tamaño de muestra necesario de cada grupo. La Tabla 1.1 ilustra el impacto tanto del tamaño de la muestra como del nivel alfa sobre la potencia. Como puede verse, la potencia llega a ser aceptable para tamaños de muestra de 100 o más en situaciones con un efecto tamaño moderado para ambos niveles de alfa. Pero cuando ocurre un efecto tamaño pequeño, las pruebas estadísticas tiene poca potencia, incluso con niveles de alfa expandidos a muestras de 200 o más. Por ejemplo, una muestra de 200 en cada grupo con un alfa de 0,05 todavía tiene un 50 por ciento de posibilidades de encontrarse diferencias significativas si el efecto tamaño es pequeño. Esto sugiere que el analista, al anticipar que los efectos van a ser pequeños, debe diseñar el estudio con muestras mucho mayores y/o niveles de alfa menos restrictivos (0,05 o 0,10). En el segundo

TABLA 1.1. Niveles de potencia para la comparación entre dos medias: variaciones por el tamaño de la muestra, el nivel de significación y el efecto tamaño

	alfa (α) = 0,05 *Efecto tamaño (ET)*		*alfa* (α) = 0,01 *Efecto tamaño (ET)*	
Tamaño muestral	*Pequeño (0,2)*	*Moderado (0,5)*	*Pequeño (0,2)*	*Moderado (0,5)*
20	0,095	0,338	0,025	0,144
40	0,143	0,598	0,045	0,349
60	0,192	0,775	0,067	0,549
80	0,242	0,882	0,092	0,709
100	0,290	0,940	0,120	0,823
150	0,411	0,990	0,201	0,959
200	0,516	0,998	0,284	0,992

Fuente: *Solo Power Analysis*, BMDP Statistical Software, Inc.

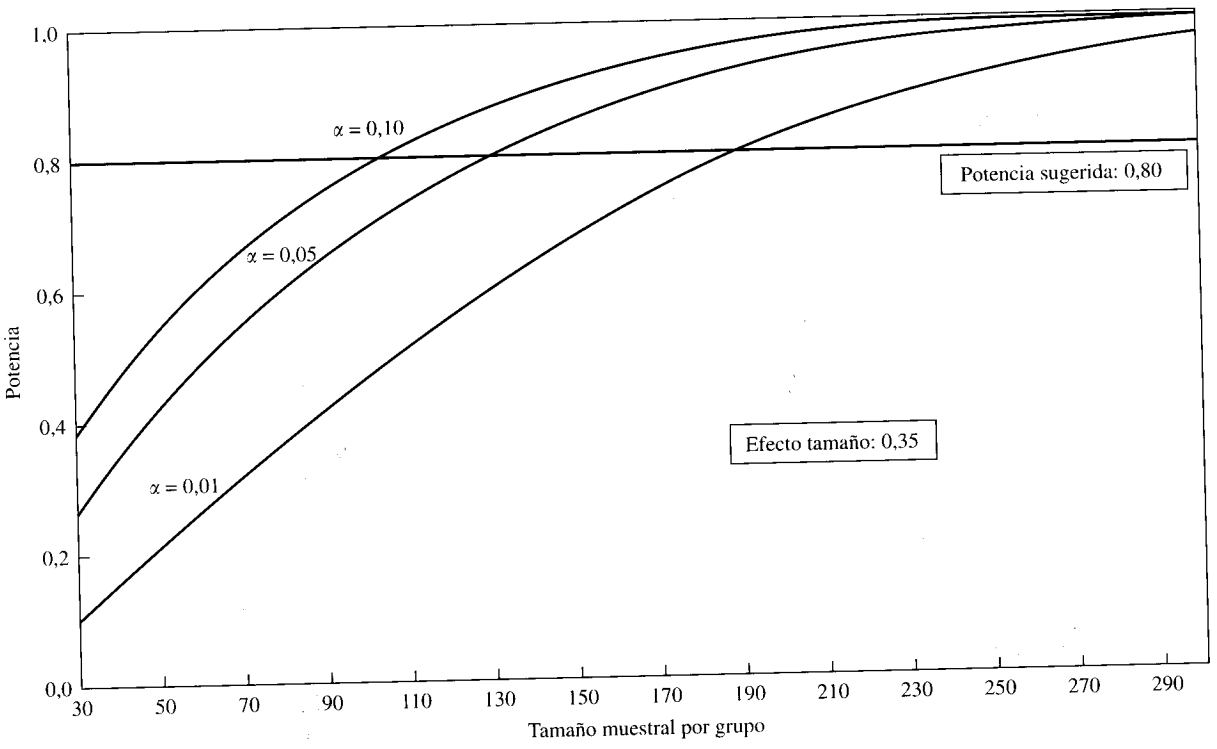


FIGURA 1.1. Impacto del tamaño de la muestra en la potencia de algunos niveles alfa (0,01; 0,05; 0,10) con efecto tamaño de 0,35.

ejemplo, la Figura 1.1 representa gráficamente la potencia para niveles de significación de 0,01; 0,5 y 0,10 con tamaños de muestra de 20 a 300 por grupo, cuando el efecto tamaño (0,35) es entre pequeño y moderado. Enfrentado a tales perspectivas, la especificación de un nivel de significación de un 0,01 requiere una muestra de 200 por grupo para conseguir el nivel deseado de potencia del 80 por ciento. Pero si se relaja el nivel alfa, se alcanza la potencia del 80 por ciento para muestras de 130 para un nivel alfa 0,05 y muestras de 100 para un nivel de significación de un 0,10.

Tales análisis permiten al investigador tomar decisiones más adecuadas en el estudio, diseño e interpretación de los resultados. Al planificar la investigación, el investigador debe estimar el efecto tamaño esperado para seleccionar entonces el tamaño de la muestra y el nivel alfa para conseguir el nivel de potencia deseado. Además de sus usos para la planificación, el análisis de potencia se utiliza también después de que el análisis ha terminado para determinar la potencia real conseguida, de tal forma que los resultados puedan ser correctamente interpretados. ¿Se deben los resultados al efecto tamaño, tamaño muestral o niveles de significación? Los analistas pueden evaluar cada uno de estos factores por su impacto sobre la significatividad o no significatividad de los resultados. El investigador puede referirse hoy en día a estudios publicados donde se analizan los detalles concretos de la determinación de la potencia [6] o acudir a varios programas de ordenador personal que asisten en los estudios de planificación para conseguir la potencia deseada o calcular la potencia de los resultados reales [2, 3]. En los Capítulos 4 y 6 se discutirán con más detalle las aplicaciones más comunes del análisis de potencia y las pautas específicas a seguir cuando se aplica regresión múltiple y el análisis multivariante de la varianza.

Habiendo ya expuesto la extensión de las técnicas multivariantes desde sus orígenes univariantes o bivariantes, introduciremos ahora brevemente cada método multivariante. A partir de la introducción de las técnicas, presentamos un esquema de clasificación para ayudar en la selección de la técnica apropiada respecto de la identificación de los objetivos de investigación (relaciones de dependencia o independencia) y el tipo de datos (métricos o no métricos).

Tipos de técnicas multivariantes

El análisis multivariante es un conjunto de técnicas de análisis de datos en expansión. Entre las técnicas más conocidas expuestas en este texto tenemos (1) regresión múltiple y correlación múltiple; (2) análisis discriminante múltiple; (3) componentes principales y análisis factorial común; (4) análisis multivariante de varianza y covarianza; (5) correlación canónica; (6) análisis cluster; (7) análisis multidimensional y (8) análisis conjunto. Entre las técnicas emergentes también incluidas están (9) análisis de correspondencias; (10) modelos de probabilidad lineal como logit y probit; y (11) modelos de ecuaciones simultáneas/estructurales. En este apartado, introduciremos cada una de las técnicas multivariantes, definiendo brevemente la técnica y el objetivo de su aplicación.

Componentes principales y análisis factorial común

El análisis factorial, que incluye variaciones tales como el análisis de componentes y el análisis factorial común, es una aproximación estadística que puede usarse para analizar interrelaciones entre un gran número de variables y explicar estas variables en términos de sus dimensiones subyacentes comunes (factores). El objetivo es encontrar un modo de condensar la información contenida en un número de variables originales en un conjunto más pequeño de variables (factores) con una pérdida mínima de información.

Si se proporciona una estimación empírica de la estructura de las variables consideradas, el análisis factorial se convierte en una base objetiva para crear escalas aditivas.

Regresión múltiple

La regresión múltiple es el método de análisis apropiado cuando el problema del investigador incluye una única variable métrica dependiente que se supone está relacionada con una o más variables métricas independientes. El objetivo del análisis de la regresión múltiple es predecir los cambios en la variable dependiente en respuesta a cambios en varias de las variables independientes. Este objetivo se consigue muy a menudo a través de la regla estadística de los mínimos cuadrados.

La regresión múltiple es útil siempre que el investigador esté interesado en predecir la cantidad o la magnitud de la variable dependiente. Por ejemplo, se puede hacer la predicción de los gastos mensuales de cenar fuera de casa (variables dependientes) con información referente a la renta familiar, su tamaño y la edad del cabeza de familia (variables independientes). De la misma forma, el investigador puede intentar predecir las ventas de una compañía a partir de información sobre sus gastos en publicidad, el número de vendedores y el número de tiendas que distribuyen sus productos.

Análisis discriminante múltiple

Si la única variable dependiente es dicotómica (es decir, comprador-no comprador) o multidicotómica (es decir, alto-medio-bajo) y por tanto no métrica, la técnica multivariante apropiada es un análisis discriminante múltiple (MDA). Como con la regresión múltiple, las variables independientes se supone que son métricas. El análisis discriminante es útil en situaciones donde la muestra total puede dividirse en grupos basándose en una variable dependiente caracterizada por varias clases conocidas. Los objetivos primarios del análisis discriminante múltiple son entender las diferencias de los grupos y predecir la verosimilitud de que una entidad (persona u objeto) pertenezca a una clase o grupo particular basándose en varias variables métricas independientes. Por ejemplo, el análisis discriminante puede usarse para distinguir innovadores de no innovadores de acuerdo a sus perfiles demográficos y psicográficos. Otras aplicaciones incluyen la distinción entre usuarios

habituales u ocasionales de un producto, compradores de marcas de ámbito nacional o restringido y el riesgo de crédito bueno del riesgo de crédito malo. Incluso la Agencia Tributaria utiliza un análisis discriminante para comparar las declaraciones seleccionadas con las devoluciones compuestas hipotéticas del contribuyente normal (para distintos niveles de renta) con el fin de identificar las devoluciones y áreas más prometedoras para la auditoría.

Análisis multivariante de la varianza y covarianza

El análisis multivariante de la varianza (MANOVA) es una técnica estadística que puede ser usada simultáneamente para explorar las relaciones entre diversas categorías de variables independientes (usualmente denominadas como **tratamientos**) y dos o más variables métricas dependientes. Como tal, representa una extensión del **análisis univariante de la varianza (ANOVA)**. El análisis multivariante de la covarianza (MANCOVA) puede usarse en conjunción con MANOVA para eliminar (después del experimento) el efecto de cualquier variable independiente no controlada sobre las variables dependientes. El procedimiento es similar al que se encuentra en la **correlación parcial bivariante**. MANOVA es útil cuando el investigador diseña una situación experimental (manipulación de varias variables de tratamiento no métricas) para comprobar hipótesis concernientes a la varianza de respuestas de grupos sobre dos o más variables métricas dependientes.

Análisis conjunto

El análisis conjunto es una técnica de dependencia emergente que ha introducido una nueva sofisticación en la evaluación de objetos, sean nuevos productos, servicios o ideas. La aplicación más directa está en productos nuevos o desarrollo de servicios, permitiendo la evaluación de productos complejos mientras que mantiene un contexto de decisión realista para el encuestado. El analista de mercado es capaz de evaluar la importancia de atributos así como los niveles de cada atributo mientras que los consumidores evalúan sólo los perfiles de unos pocos productos, que son combinaciones de niveles de producto. Por ejemplo, un concepto de un producto que tiene tres atributos (precio, calidad y color), cada uno de los cuales a tres niveles (por ejemplo, rojo, amarillo y azul). En lugar de tener que evaluar todas las 27 combinaciones posibles ($3 \times 3 \times 3$), se puede evaluar un subconjunto (9 o más) por su atractivo para los consumidores, y el investigador sabe no sólo cuál es la importancia de cada atributo sino también la importancia de cada nivel (el atractivo del rojo frente al amarillo y frente al azul). Más aún, cuando se completan las evaluaciones del consumidor, pueden usarse los resultados del análisis conjunto en simuladores del diseño del producto, que mostrarán la aceptación del cliente para cualquier número de formulaciones de producto y ayudar en el diseño del producto óptimo.

Correlación canónica

El análisis de correlación canónica puede verse como una extensión lógica de un análisis de regresión múltiple. Recordemos que el análisis de regresión múltiple implica una única variable dependiente métrica y varias variables métricas independientes. Con el análisis canónico el objetivo es correlacionar simultáneamente varias variables dependientes métricas y varias variables métricas independientes. Mientras que la regresión múltiple implica una única variable dependiente, la correlación canónica implica múltiples variables dependientes. El principio subyacente es desarrollar una combinación lineal de cada conjunto de variables (tanto independientes como dependientes) para maximizar la correlación entre los dos conjuntos. O dicho de otra forma, el procedimiento implica obtener un conjunto de ponderaciones para las variables dependientes e independientes que proporcione la correlación única máxima entre el conjunto de variables dependientes y el conjunto de variables independientes.

Análisis cluster

El análisis cluster es una técnica analítica para desarrollar subgrupos significativos de individuos u objetos. De forma específica, el objetivo es clasificar una muestra de entidades (personas u objetos) en un número pequeño de grupos mutuamente excluyentes basados en similitudes entre las entidades. En el análisis cluster, a diferencia del análisis discriminante, los grupos no están predefinidos. Por consiguiente, se usa la técnica para identificar los grupos.

Habitualmente, el análisis cluster implica al menos dos etapas. La primera es la medida de alguna forma de similitud o asociación entre las entidades para determinar cuántos grupos existen en realidad en la muestra. La segunda etapa es describir las personas o variables para determinar su composición. Este paso puede llevarse a cabo aplicando el análisis discriminante a los grupos identificados por la técnica cluster.

Análisis multidimensional

En el análisis multidimensional, el objetivo es transformar los juicios de los consumidores de similitud o preferencia (por ejemplo, preferencias por tiendas o marcas comerciales) en distancias representadas en un espacio multidimensional. Si los objetos A y B son en opinión de los encuestados más similares que el resto de los pares posibles de objetos, las técnicas de análisis multidimensional situarán a los objetos A y B de tal forma que la distancia entre ellos en un espacio multidimensional es menor que la distancia entre cualquier otro par de objetos. Los mapas perceptuales resultantes muestran el posicionamiento relativo entre losobjetos, pero es necesario un análisis adicional para evaluar qué atributos predicen la posición de cada objeto.

Análisis de correspondencias

Para finalizar, el análisis de correspondencias es una técnica de interdependencia recientemente desarrollada que facilita tanto la reducción dimensional de una clasificación de objetos (por ejemplo, productos, personas, etc.,) sobre un conjunto de atributos y el mapa perceptual de objetos relativos al estos atributos. Los investigadores se enfrentan constantemente a la necesidad de «cuantificar datos cualitativos» que encuentran en variables nominales. El análisis de correspondencias difiere de otras técnicas de interdependencia discutidas antes en su capacidad para acomodar tanto datos no métricos como relaciones no lineales.

En su forma más básica, el análisis de correspondencias emplea una tabla de contingencia, que es la tabulación cruzada de dos variables categóricas. A continuación transforma los datos no métricos en un nivel métrico y realiza una reducción dimensional (similar al análisis factorial) y un mapa perceptual (similar al análisis multidimensional). A modo de ejemplo, las preferencias por una marca de los encuestados pueden ser tabuladas de forma cruzada con variables demográficas (por ejemplo, género, categorías de renta, ocupación) indicando cuánta gente que prefiere cada una de las marcas entra dentro de cada categoría de las variables demográficas. A través del análisis de correspondencias, la asociación o «correspondencia» de marcas y las características distintivas de aquellos que prefieren cada marca se muestran en un mapa bi o tridimensional, tanto de marcas como características de los encuestados. Las marcas percibidas como similares están localizadas en una cercana proximidad unas de otras. De la misma forma, las características más distintivas de los encuestados que prefieren cada marca están determinadas también por la proximidad de las categorías de las variables demográficas respecto de la posición de la marca. El análisis de las correspondencias proporciona una representación multivariante de la interdependencia de datos no métricos que no es posible realizar con otros métodos.

Modelos de probabilidad lineal

Los modelos de probabilidad lineal, a menudo llamados *análisis logit*, consisten en una combinación de regresión múltiple y análisis de discriminante múltiple. Esta técnica es similar al análisis de regresión múltiple en que una o más variables independientes se usan para predecir una única variable dependiente. Lo que distingue un modelo de probabilidad lineal de la regresión múltiple es que la variable dependiente es no métrica, como en el análisis discriminante. La escala no métrica de la variable dependiente requiere diferencias en el método de estimación y supuestos sobre el tipo de distribución subyacente, siendo en la mayoría de sus otras facetas similar a la regresión múltiple. Por tanto, una vez que la variable dependiente está especificada correctamente y se emplea la técnica de estimación apropiada, se usan igualmente los supuestos básicos considerados en la regresión múltiple. Los modelos de probabilidad lineal se distinguen del análisis discriminante en que acomodan todos los tipos de variables independientes (métricas y no métricas) y no requieren el supuesto de normalidad multivariante. Sin embargo, en muchos casos, particularmente con más de dos niveles de la variable dependiente, el análisis discriminante es la técnica más apropiada.

Modelos de ecuaciones estructurales

El modelo de ecuaciones estructurales, a menudo denominado simplemente como LISREL (el nombre de uno de los paquetes informáticos más populares), es una técnica que permite separar las relaciones para cada conjunto de variables dependientes. En su acepción más simple, el modelo de ecuaciones estructurales proporciona la técnica de estimación más adecuada y eficiente para series de estimaciones de ecuaciones simultáneas mediante regresiones múltiples. Se caracteriza por dos componentes básicos: (1) el modelo estructural y (2) el modelo de medida. El *modelo estructural* es el modelo «guía», que relaciona variables independientes y variables dependientes. En tales situaciones, la teoría, antes que la experiencia u otras directrices, permitirá al investigador distinguir qué variables independientes predicen cada variable dependiente. Los modelos previamente discutidos que incluyen múltiples variables dependientes —análisis multivariante de la varianza y correlación canónica— no son apropiados en esta situación, dado que permiten sólo una *única* relación entre variables dependientes e independientes.

El *modelo de medida* permite al investigador usar varias variables (**indicadores**), para una única variable dependiente o independiente. Por ejemplo, la variable dependiente puede ser un concepto representado por una escala aditiva, tal como el amor propio. En el modelo de medida el investigador puede evaluar la contribución de cada ítem de la escala así como incorporar cómo la escala mide el concepto (fiabilidad) en la estimación de las variables dependientes e independientes. Este procedimiento es similar al desarrollo del análisis factorial (discutido en una sección posterior) de los ítems de la escala y utiliza las cargas factoriales en la regresión.

Otras técnicas multivariantes emergentes

El uso generalizado de las herramientas informáticas ayudó a iniciar la era del análisis multivariante tal y como lo conocemos hoy, con un número de técnicas especializadas que se pueden aplicar a una gama amplia de situaciones. No obstante, ahora nos encontramos al principio de una era en la cual el análisis multivariante incorpora nuevos enfoques para identificar y representar las relaciones multivariantes. Un área de desarrollo en el análisis multivariante es la búsqueda de datos y las redes neuronales. La búsqueda de datos es el intento de cuantificar las relaciones entre grandes cantidades de información con una especificación previa mínima de la naturaleza de las relaciones. Una técnica que se usa muchas veces junto con la búsqueda de datos son las redes neuronales, una técnica de análisis flexible que es capaz de llevar a cabo una identificación

de relaciones (parecida a la regresión múltiple o al análisis discriminante) o la reducción de datos y el análisis estructural (semejante al análisis factorial o cluster). Las redes neuronales son diferentes a las técnicas multivariantes más tradicionales citadas previamente tanto en la formulación del modelo como en los tipos de relaciones más complejos que se pueden formular. Otra área que incluye un distanciamiento de la teoría estadística inferencial tradicional es el desarrollo de la técnica de la muestra repetida o «arranque». Esta técnica elimina la necesidad de cumplir determinados supuestos estadísticos (como la normalidad), mediante el uso del ordenador para replicar una «muestra repetida» de la muestra original, con el reemplazo y la generación de una estimación empírica de la distribución muestral. Se puede encontrar un resumen de estas nuevas técnicas en el Capítulo 12.

Una clasificación de las técnicas multivariantes

Para ayudarle a familiarizarse con las técnicas multivariantes, presentamos una clasificación de los métodos multivariantes en la Figura 1.2. Esta clasificación se basa en tres juicios que el analista debe hacer sobre el objeto a investigar y la naturaleza de los datos: (1) ¿pueden dividirse las variables en dependientes o independientes basándose la clasificación en alguna teoría? (2) Si puede hacerse, ¿cuántas de estas variables son tratadas como dependientes en un análisis simple? (3) ¿Cómo son las variables medidas? La selección de la técnica multivariante apropiada depende de las respuestas a estas tres cuestiones.

Cuando consideramos la aplicación de las técnicas estadísticas multivariantes, la primera cuestión que nos debemos preguntar es, ¿pueden dividirse las variables mediante la clasificación de dependiente e independiente? La respuesta a esta cuestión indica si se debería utilizar un análisis de dependencia o interdependencia. Nótese que en la Figura 1.2, las técnicas de dependencia están en el lado izquierdo y las técnicas de interdependencia están a la derecha. Un **análisis de dependencia** puede definirse como aquel en el que una variable o conjunto de variables es identificado como la **variable dependiente** y que va a ser explicada por otras variables conocidas como **variables independientes**. Como ejemplo de una dependencia técnica tenemos el análisis de regresión múltiple. Como contraste, un análisis de **interdependencia** es aquel en que ninguna variable o grupo de variables es definido como independiente o dependiente. Más bien, el procedimiento implica el análisis de todas las variables del conjunto simultáneamente. El análisis factorial es un ejemplo de un análisis de interdependencia. Pero centrémonos en el análisis de dependencia primero y usemos la clasificación de la Figura 1.2 para seleccionar el método multivariante apropiado.

Los diferentes métodos que constituyen el análisis de dependencia pueden ser a su vez divididos en dos tipos según: (1) el número de variables dependientes y (2) el tipo de escalas de medida empleadas para las variables. Teniendo en cuenta el número de variables dependientes, el análisis de dependencia puede clasificarse como aquel que tiene tanto una variable dependiente única como varias variables dependientes o incluso varias relaciones de dependencia/independencia. El análisis de dependencia puede incluso ser clasificado en función del tipo de escala de la variable con variables métricas (numéricas/cuantitativas) o no métricas (cualitativas/categóricas). Si el análisis implica una única variable dependiente que es métrica, la técnica apropiada es tanto el análisis de regresión múltiple como el análisis conjunto. El análisis conjunto es un caso especial. Se trata de un procedimiento de dependencia que puede tratar la variable dependiente como métrica o no métrica, en función de las circunstancias. Por otro lado, si la única variable dependiente es no métrica (categórica), entonces la técnica apropiada es, o bien el análisis discriminante múltiple, o bien los modelos de probabilidad lineal. En contraste, cuando el problema del investigador implica varias variables dependientes, hay otras cuatro técnicas estadísticas apropiadas. Si varias variables dependientes son métricas, debemos entonces mirar a las variables independientes. Si las variables independientes son no métricas, debemos elegir la técnica multivariante de análisis de la varianza. Si las variables independientes son métricas, la apropiada es la correlación canónica. Si

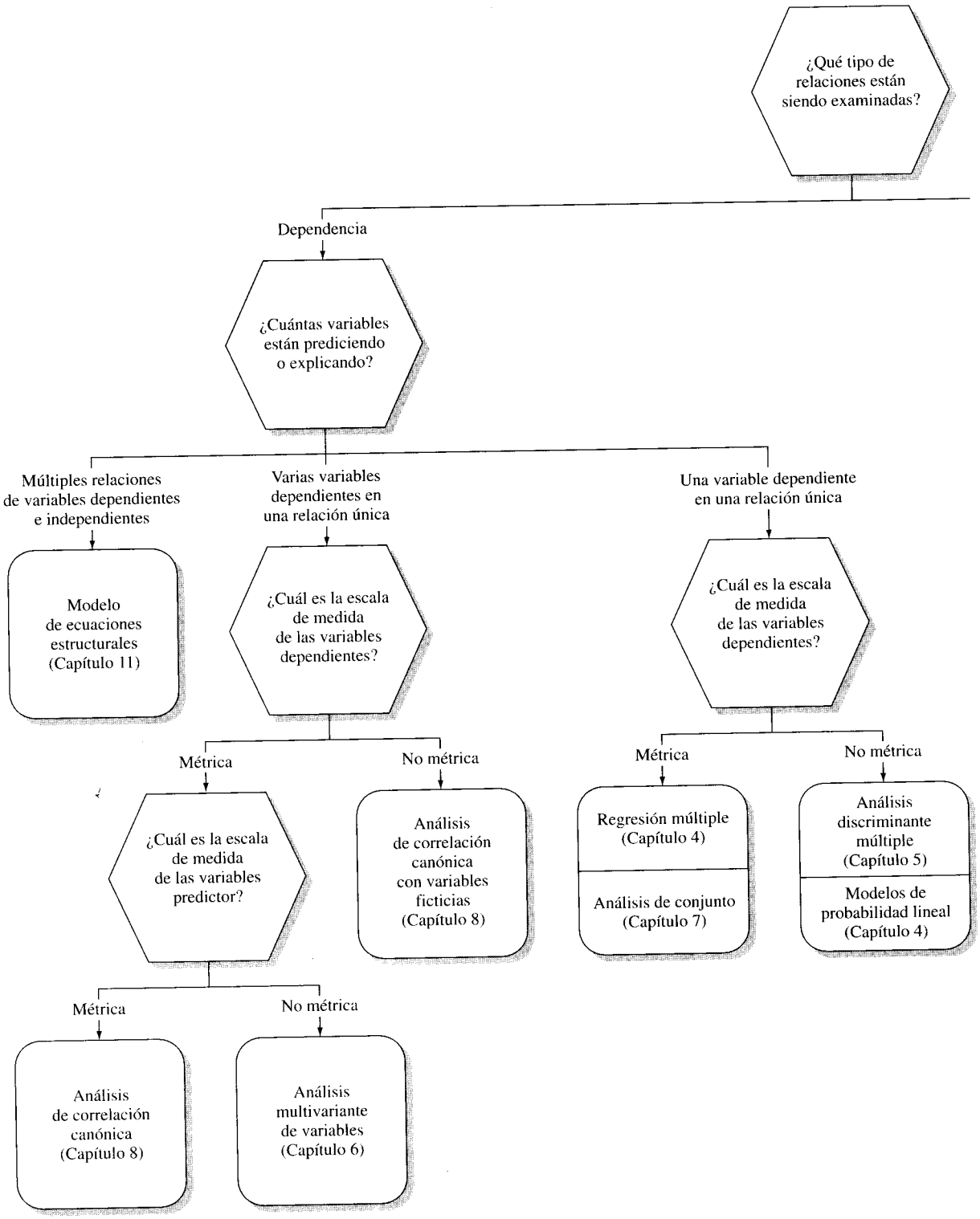
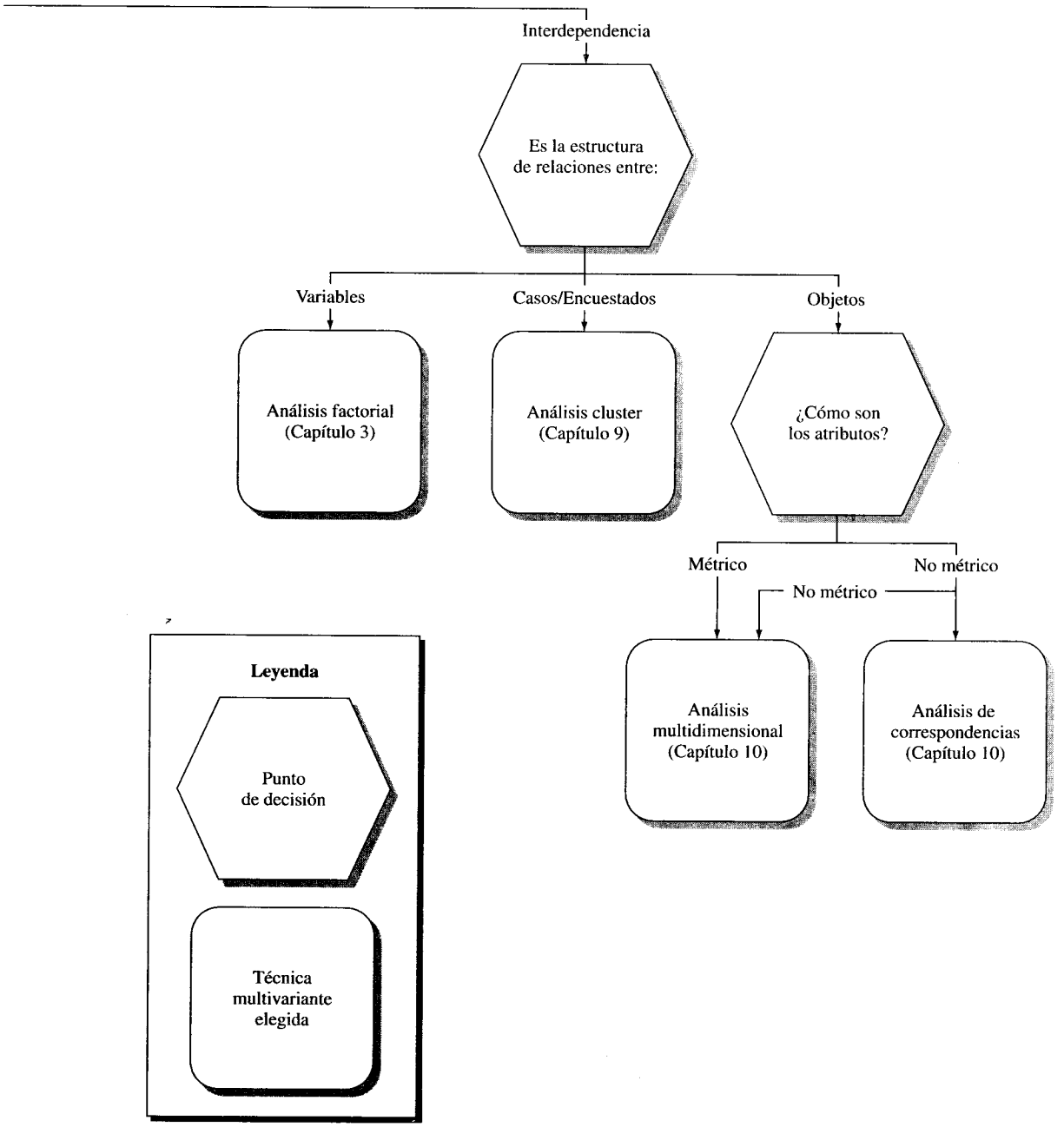


FIGURA 1.2. Selección de una técnica multivariante.


Interdependencia
Es la estructura de relaciones entre:
Variables
Casos/Encuestados
Objetos
Análisis factorial (Capítulo 3)
Análisis cluster (Capítulo 9)
¿Cómo son los atributos?
Métrico
No métrico
No métrico
Análisis multidimensional (Capítulo 10)
Análisis de correspondencias (Capítulo 10)
Leyenda
Punto de decisión
Técnica multivariante elegida

varias variables dependientes son no métricas, entonces pueden transformarse a través de una variable ficticia de código (0-1) y puede utilizarse también el análisis canónico*. Finalmente, si se postula un conjunto de relaciones de variables dependientes/independientes, entonces el modelo de ecuaciones estructurales es el apropiado.

Existe una estrecha relación entre los diversos procedimientos de dependencia, que pueden ser vistos como una familia de técnicas. La Tabla 1.2 define varios de los análisis de dependencia multivariante en términos de la naturaleza y número de las variables dependientes e independientes. Como podemos ver, la correlación canónica puede considerarse como el modelo general en el cual se basan otras muchas técnicas multivariantes, dado que sitúa la mínima restricción respecto al tipo y número de variables tanto de valor teórico dependiente como independiente. Como las restricciones están basadas en valores teóricos, pueden alcanzarse conclusiones más precisas apoyándonos en la escala específica empleada en la medición de los datos. Tales técnicas multivariantes van desde el método general del análisis canónico al más especializado método de modelización de ecuaciones estructurales.

El análisis de interdependencia se muestra en el lado derecho de la Figura 1.2. Los lectores recordarán que en el análisis de interdependencia las variables no puede ser clasificadas como

TABLA 1.2. Las relaciones entre métodos de dependencia multivariante

Correlación canónica

$$Y_1 + Y_2 + Y_3 + \ldots + Y_n = X_1 + X_2 + X_3 + \ldots + X_n$$

(*métrica, no métrica*) (*métrica, no métrica*)

Análisis multivariante de la varianza

$$Y_1 + Y_2 + Y_3 + \ldots + Y_n = X_1 + X_2 + X_3 + \ldots + X_n$$

(*métrica*) (*no métrica*)

Análisis de la varianza

$$Y_1 = X_1 + X_2 + X_3 + \ldots + X_n$$

(*métrica*) (*no métrica*)

Análisis discriminante múltiple

$$Y_1 = X_1 + X_2 + X_3 + \ldots + X_n$$

(*no métrica*) (*métrica*)

Análisis de regresión múltiple

$$Y_1 = X_1 + X_2 + X_3 + \ldots + X_n$$

(*métrica*) (*métrica, no métrica*)

Análisis conjunto

$$Y_1 = X_1 + X_2 + X_3 + \ldots + X_n$$

(*no métrica, métrica*) (*no métrica*)

Modelo de ecuaciones estructurales

$$Y_1 = X_{11} + X_{12} + X_{13} + \ldots + X_{1n}$$
$$Y_2 = X_{21} + X_{22} + X_{23} + \ldots + X_{2n}$$
$$Y_m = X_{m1} + X_{m2} + X_{m3} + \ldots + X_{mn}$$

(*métrica*) (*métrica, no métrica*)

* Las **variables ficticias** se tratan con más detalle más adelante. En resumen, la codificación de la variable ficticia es una manera de transformar datos no métricos en datos métricos. Incluye la creación de las denominadas variables ficticias, en las cuales se asignan unos y ceros al sujeto, dependiendo de si cuenta o no con cierta característica. Por ejemplo, si un sujeto es masculino se le asigna un 0 y si el sujeto es femenino se le asigna un 1, o al contrario.

dependientes o independientes. En vez de esto, todas las variables son analizadas simultáneamente en un esfuerzo por encontrar una estructura subyacente para el conjunto total de variables o sujetos. Si se está analizando la estructura de las variables, entonces el análisis factorial es la técnica apropiada. Si los casos o los encuestados se van a agrupar para representar una estructura, entonces seleccionaremos el análisis cluster. Finalmente, si el interés está en la estructura de objetos, deberían aplicarse las técnicas de análisis multidimensional. Como ocurre con el análisis de dependencia, deberían considerarse las propiedades de las técnicas de medición. Generalmente, el análisis factorial y el análisis cluster se consideran análisis de interdependencia métricos. Sin embargo, los datos no métricos pueden ser transformados a través de una variable ficticia codificada para usarlos con análisis factorial y análisis cluster. Se han desarrollado tanto las aproximaciones métricas como las no métricas al análisis multidimensional. Si se van a analizar las interdependencias entre objetos medidos por datos no métricos, el análisis de correspondencias es la técnica apropiada.

Orientaciones para el análisis multivariante y su interpretación

Como se ha podido comprobar, el análisis multivariante tiene un carácter variado y puede ser bastante poderoso. Este poder es especialmente tentador cuando el investigador no está seguro del diseño del análisis más apropiado y utiliza el análisis multivariante como un sustituto del necesario análisis conceptual. Incluso cuando se aplica correctamente, los esfuerzos por acomodar las múltiples variables y relaciones crean complejidades adicionales en los resultados y su interpretación. Por tanto, advertimos contra su uso sin la base conceptual apropiada para apoyar la técnica seleccionada sobre aquellos conceptos básicos mencionados previamente y los temas abordados en la siguiente sección.

Hemos discutido también varios asuntos particularmente aplicables al análisis multivariantes. Por tanto, mientras no exista una única «respuesta», hemos encontrado que el análisis y la interpretación de cualquier problema multivariante puede verse ayudado por un conjunto general de directrices. No se trata de ningún modo de una lista exhaustiva de consideraciones, sino que la lista representa más bien una «filosofía del análisis multivariante». Las siguientes secciones discuten estos puntos pero no en un orden concreto, sino haciendo igual énfasis en todos ellos.

Establecer la significación práctica así como la estadística

La fuerza del análisis multivariante reside en sus medios aparentemente «mágicos» para clasificar una variedad de posibles alternativas y encontrar aquellas que tienen significación estadística. Pero con este poder debemos tener precaución. Muchos investigadores se vuelven miopes al fijarse solamente en la significación conseguida por los resultados sin entender sus interpretaciones, buenas o malas. En su lugar, el investigador debe atender no sólo a la significación estadística de los resultados sino también a su significación práctica. La **significación práctica** se refiere a la cuestión, «¿y para qué?». Para cualquier aplicación en la gestión, los resultados deben tener un efecto demostrable que justifique la acción. En el terreno académico, el investigador se llega a fijar no sólo en la significación estadística de los resultados sino también en sus implicaciones teóricas y sustantivas, que en muchas ocasiones se deducen de su significación práctica.

Como ejemplo ilustrativo de esta situación consideramos un análisis de regresión para predecir las intenciones de compra, medidas como la probabilidad entre 0 y 100 de que el cliente volverá a comprar a la empresa. El estudio se lleva a cabo y el resultado es significativo al nivel de significación de 0,05. Los ejecutivos aceptan los resultados y modifican la estrategia de la empresa. Pero lo que no se ha percibido es que mientras la relación era significativa, la capacidad predictiva era baja, tan baja que la estimación de la posibilidad de repetir compra podría variar tanto como

un 20 por ciento al nivel de significación del 0,05. ¡La relación de la «significación estadística» podría entonces tener un rango de error de 40 puntos porcentuales! Un cliente del cual se predice que tiene una oportunidad de volver de 50/50 podría realmente tener probabilidades del 30 al 70 por ciento, representando niveles inaceptables sobre los cuales actuar. Los investigadores y los gerentes no han probado la significación práctica o de gestión de los resultados, olvidando que la relación todavía necesitaba un ulterior refinamiento.

El tamaño muestral afecta a todos los resultados

La discusión de la potencia estadística demuestra que el impacto sustancial del tamaño muestral opera en la consecución de la significación estadística, tanto en tamaños muestrales grandes como pequeños. Para muestras pequeñas, la sofistificación y la complejidad del análisis multivariante puede fácilmente resultar tanto en (1) muy poca potencia estadística de la prueba para identificar de forma realista resultados significativos o (2) fácilmente un «sobreaprovechamiento» de los datos de tal forma que sean artificialmente buenos porque se ajustan muy bien a la muestra, aunque no sean generalizables. Lo mismo ocurre para muestras grandes que, como ya se ha discutido antes, pueden hacer a los test estadísticos altamente sensibles. Siempre que los tamaños muestrales excedan los 200 o 400 encuestados, el investigador debería examinar todos los resultados significativos para asegurarse que tienen significación práctica debido al aumento de la potencia estadística como consecuencia del tamaño muestral. Los tamaños muestrales también afectan a los resultados cuando los análisis implican grupos de encuestados, como ocurre en el análisis discriminante o en MANOVA. Tamaños muestrales desiguales entre los grupos influencian a los resultados y requieren un análisis y/o interpretación adicional. Por tanto, el investigador o usuario del análisis multivariante debería siempre valorar los resultados a la luz de la muestra utilizada.

Conocer los datos

Las técnicas del análisis multivariante, por su propia naturaleza, identifican relaciones complejas que son difíciles de representar de forma simple. Como resultado, la tendencia es aceptar los resultados sin el típico examen que uno emprende en los análisis univariante y bivariante (por ejemplo, gráfico de dispersión de correlaciones y boxplots de comparaciones de media). Pero estos «atajos» pueden ser el preludio del desastre. El análisis multivariante requiere un examen incluso *más riguroso* de los datos porque la influencia de atípicos, violaciones de los supuestos y la pérdida de datos puede agravarse a través de varias variables y tener efectos sustancialmente diferentes. Para servirse de todos los beneficios de las técnicas multivariantes, el analista debe también «saber dónde mirar» con formulaciones alternativas del modelo original, tales como relaciones no lineales e interactivas. El analista tiene, sin embargo, un conjunto de técnicas de diagnóstico en continua expansión que permiten que estas relaciones multivariantes sean descubiertas por medios similares a los métodos univariantes y bivariantes. El investigador de un problema multivariante debe tomarse su tiempo en utilizar estas medidas de diagnóstico para un mayor entendimiento de los datos y de las relaciones básicas que existen.

Procurar la parsimonia del modelo

Las técnicas multivariantes se diseñan para acomodar las variables en el análisis. Este carácter, sin embargo, no debería sustituir el desarrollo de modelos conceptuales *antes* de que se apliquen las técnicas multivariantes. Aunque es siempre importante evitar omitir una variable predictor crítica, denominada **error de especificación**, por varias razones el analista debe también intentar evitar insertar variables indiscriminadamente. En primer lugar, las variables irrelevantes habitualmente aumentan la capacidad del análisis para ajustar la muestra de datos pero a costa de sobreajustar los datos y hacerlos menos generalizables para la población. En segundo lugar, las variables irrelevantes

no sesgan típicamente las estimaciones de las variables relevantes, pero pueden enmascarar los efectos verdaderos debido a la multicolineadad. La **multicolinealidad** representa el grado en el que cualquier efecto de una variable puede ser prevista o explicada por las otras variables del análisis. A medida que aumenta la multicolinealidad, la capacidad para definir el efecto de cualquier variable disminuye. Por tanto, incluyendo variables que no son relevantes conceptualmente podemos tener varios efectos potencialmente dañinos, incluso si las variable adicionales no sesgan directamente los resultados del modelo.

Atender a los errores

Incluso con la capacidad del análisis multivariante, difícilmente conseguiremos la mejor predicción en el primer análisis. El analista se enfrenta con la cuestión, «¿adónde podemos ir desde aquí?». La mejor respuesta es mirar a los errores en la predicción, tanto si son los residuos del análisis de regresión, la ausencia de clasificación de observaciones en el análisis discriminante o los atípicos del análisis cluster.

En cada caso, el analista debería utilizar los errores de predicción no como una medida de error o como algo meramente a eliminar, sino como un punto de partida para diagnosticar la validez de los resultados obtenidos y como una indicación de las relaciones que quedan sin explicar.

Validar los resultados

La capacidad del análisis multivariante para identificar interrelaciones complejas también implica que puede darse el caso de que los resultados sean específicos sólo para la muestra y no generalizables a la población. El investigador debe siempre asegurar que existen observaciones suficientes por parámetro estimado para evitar el «sobreajuste» de la muestra, como se ha discutido antes. Pero igual de importantes son los esfuerzos destinados a validar los resultados mediante diferentes métodos, que incluyen (1) división de la muestra y el uso de una submuestra para estimar el modelo y usar una segunda submuestra para estimar la precisión predictiva, (2) empleo de un análisis de «bootstrapping» [9], o (3) incluso conseguir una muestra distinta para asegurar que los resultados son apropiados para otras muestras. Cualquiera que sea la técnica multivariante empleada, el investigador debe centrarse no sólo en estimar un modelo significativo sino también en asegurar que es representativo de la población en su conjunto. Recordemos que el objetivo no es encontrar el mejor «ajuste» sólo para la muestra sino desarrollar el modelo que mejor describa a la población en su conjunto.

Una aproximación organizada al diseño de modelos multivariantes

Al discutir las numerosas técnicas multivariantes a disposición del investigador y la gran cantidad de supuestos que implica su aplicación, se hace evidente que finalizar con éxito un análisis multivariante implica algo más que la selección del método correcto. Deben resolverse asuntos que van desde el problema de definición al diagnóstico crítico de los resultados. Para ayudar al investigador o usuario en la aplicación de los métodos multivariantes, presentamos una aproximación al análisis multivariante en seis pasos. La intención no es proporcionar un conjunto rígido de procedimientos a seguir, sino más bien ofrecer una serie de pautas que facilitan la aproximación al diseño de modelos. La aproximación a la modelización se centra en el análisis de un plan de investigación bien definido, comenzando con un modelo conceptual que detalle las relaciones a examinar. Una vez definido en términos conceptuales, se pueden iniciar los trabajos empíricos, incluyendo

la selección de una técnica multivariante específica y su puesta en práctica. Después de haber obtenido los resultados significativos, su interpretación se convierte en el asunto central, dirigiendo una atención especial al valor teórico. Finalmente, la medidas de diagnóstico aseguran que el modelo no sólo es válido para la muestra de datos sino que es también generalizable. La siguiente exposición describe resumidamente cada paso de esta aproximación.

El proceso de modelización en seis pasos proporciona un esquema para desarrollar, interpretar y validar cualquier análisis multivariante. Cada investigador debe desarrollar su propio criterio de «éxito o fracaso» en cada nivel, pero las discusiones de cada técnica proporcionan pautas siempre aplicables. Nuestro deseo es que el énfasis en la aproximación al diseño de modelos proporcione una base amplia para la interpretación, estimación y desarrollo de las técnicas que mejorarán los análisis multivariantes del usuario.

Primer paso: Definir el problema de investigación, objetivos y técnica multivariante conveniente

El punto de partida para cualquier análisis multivariante es definir el problema de investigación y los objetivos analíticos en términos conceptuales, antes de especificar cualquier variable o medida. No se puede exagerar el papel del desarrollo del modelo conceptual o teoría. No importa si la investigación es práctica o académica, el investigador debe ver en primer lugar el problema en términos conceptuales, definiendo los conceptos e identificando las relaciones fundamentales a investigar. Desarrollar un modelo conceptual no es dominio exclusivo de los académicos; también se ajusta a la aplicación a la experiencia del mundo real.

Un modelo conceptual no necesita ser complejo y detallado, sino una simple representación de relaciones a estudiar. Si se propone una relación de dependencia como objetivo de investigación, el investigador debe especificar los conceptos dependientes e independientes. Nótese que se define un concepto, más que una variable. Para la aplicación de una técnica de interdependencia, se deberían determinar las dimensiones de la estructura o similitud. Tanto en situaciones de dependencia como de interdependencia, el investigador identifica primero las ideas o temas de interés en lugar de fijarse en las medidas a utilizar. Con ello se minimiza la posibilidad de que conceptos relevantes sean omitidos en el esfuerzo por desarrollar medidas y definir los detalles del diseño. Remitimos a los lectores interesados en el desarrollo de modelos conceptuales al Capítulo 11.

Con los objetivos y el modelo conceptual especificados, el analista sólo tiene que elegir la técnica multivariante apropiada. Después de haber seleccionado un método de dependencia o interdependencia, la última decisión consiste en elegir la técnica específica basada en las características de medición de las variables dependientes e independientes. Se pueden especificar las variables antes del estudio; se pueden definir después de recoger los datos.

Segundo paso: Desarrollo del proyecto de análisis

Con el modelo conceptual establecido, la atención se dirige a la puesta en práctica de la técnica elegida. Para cada técnica, el investigador debe desarrollar un plan de análisis específico que dirija el conjunto de supuestos que subyacen en la aplicación de la técnica. Los supuestos van desde consideraciones generales de tamaños de muestra mínimos o deseados, pasando por los tipos de variables permitidas o requeridas (métricas versus no métricas) y métodos de estimación, a supuestos tales como el tipo de medida de asociación usada en el análisis multidimensional, la estimación de los resultados agregados o desagregados en el conjunto o el uso de formulaciones especiales de variables para representar efectos interactivos o no lineales en la regresión. En cada caso, estos supuestos resuelven los detalles específicos y finalizan la formulación del modelo y los requisitos del esfuerzo de recogida de datos.

Tercer paso: Evaluación de los supuestos básicos de la técnica multivariante

Con la recogida de datos, el primer análisis no consiste en estimar el modelo multivariante, sino en evaluar los supuestos subyacentes. Todas las técnicas multivariantes tienen supuestos subyacentes, tanto estadísticos como conceptuales, que afectan sustancialmente a su capacidad para representar relaciones multivariantes. Para las técnicas basadas en la inferencia estadística se deben tener en cuenta los supuestos de normalidad multivariante, linealidad, independencia de los términos de error e igualdad de las varianzas en una relación de dependencia. En el Capítulo 2 se discute sobre la valoración de estos supuestos con más detalle. Cada técnica tiene una serie de supuestos conceptuales que tratan sobre asuntos como la formulación de modelos y los tipos de representaciones. Antes de intentar cualquier estimación del modelo, el investigador debe asegurarse que se encuentran cumplidos tanto los supuestos estadísticos como los conceptuales.

Cuarto paso: Estimación del modelo multivariante y valoración del ajuste del modelo

Una vez satisfechos los modelos, se procede a la estimación efectiva del modelo multivariante y a una valoración global del ajuste del modelo. En el proceso de estimación, el analista puede optar por distintas opciones para elegir las características específicas de los datos (por ejemplo, uso de covarianzas en MANOVA) o maximizar el ajuste de los datos (por ejemplo, rotación de los factores o funciones discriminantes). Después de haber estimado el modelo, se evalúa el ajuste para averiguar si consiguen niveles aceptables sobre los criterios estadísticos (por ejemplo, nivel de significación), identifica las relaciones propuestas y consigue la significación práctica. En muchas ocasiones, el modelo se volverá a especificar en un intento de conseguir mejorar los niveles de ajuste y/o explicación.

Con independencia del nivel de ajuste que se encuentre, el analista debe también determinar si los resultados están excesivamente afectados por un único o pequeño conjunto de observaciones que indican que los resultados pueden ser inestables. Estos esfuerzos aseguran que los resultados son «robustos» y estables al aplicarlos razonablemente a todas las observaciones de la muestra. Las observaciones mal ajustadas pueden identificarse como atípicas, observaciones influyentes u otros resultados dispersos (por ejemplo, conglomerados de un único miembro o casos seriamente desclasificados en el análisis discriminante).

Quinto paso: Interpretar el valor teórico

Con un nivel aceptable de ajuste del modelo, al interpretar el valor o valores teóricos se revela la naturaleza de las relaciones multivariantes. La interpretación de los efectos para variables individuales se realiza examinando los coeficientes estimados (ponderaciones) para cada variable en el valor teórico (por ejemplo, ponderaciones de regresión, cargas de los factores o utilidades conjuntas). Más aún, algunas técnicas también estiman los valores teóricos múltiples que representan las dimensiones subyacentes de la comparación o asociación (por ejemplo, funciones discriminantes o componentes principales). La interpretación puede conducir a re-especificaciones adicionales de las variables y/o formulación del modelo, donde el modelo se estima de nuevo y se interpreta una vez más. El objetivo es identificar la evidencia empírica de las relaciones multivariantes de los datos muestrales que pueden generalizarse para el total de la población.

Sexto paso: Validación del modelo multivariante

Antes de aceptar los resultados, el investigador debe someterlos a un conjunto final de diagnósticos que aseguran el grado de generalidad de los resultados por los métodos de validación disponibles. Los intentos de validar el modelo se dirigen directamente hacia la demostración de la generalidad de los resultados al conjunto de la población (véase la anterior discusión de las técnicas de validación). Ambos diagnósticos añaden poco a la interpretación de los resultados pero sirven para asegurar los resultados más descriptivos de los datos y su generalización al conjunto de la población.

Un diagrama de flujos de decisiones

Para cada técnica multivariante, el uso de los seis pasos de construcción de un modelo multivariante se indicará en un diagrama de flujos de decisiones segregado en dos secciones. La primera sección (los pasos 1 a 3) se refiere a los temas abordados con la preparación para la propia estimación de modelos (es decir, los objetivos de investigación, consideraciones para el diseño de la investigación y el ensayo para las suposiciones). La segunda sección del diagrama de flujos de decisiones (los pasos 4 a 6) se refiere a las cuestiones pertinentes a un modelo de estimación, interpretación y validación. El diagrama de flujos de decisiones proporciona al investigador un método simplificado pero sistemático para la aplicación de la aproximación organizada al diseño de modelos multivariantes cuando se aplica cualquier técnica multivariante.

Bases de datos

Para explicar e ilustrar cada una de las técnicas multivariantes de forma más extensa, a lo largo del libro utilizaremos conjuntos de datos hipotéticos. Estos conjuntos de datos se obtuvieron todos de la Compañía Hair, Anderson y Tatham (HATCO), un enorme (aunque inexistente) distribuidor industrial. Cada una de las series de datos se obtuvo de las encuestas de clientes de HATCO, realizadas por una empresa de investigación del mercado reconocida.

La principal base de datos

La principal base de datos, consistente en 100 observaciones de 14 variables separadas, es un ejemplo de un estudio de segmentación de la situación empresa a empresa, específicamente un informe sobre los clientes actuales de HATCO. Se utilizan tres tipos de datos. La primera clase es la percepción de HATCO sobre siete atributos identificados en estudios pasados como los más influyentes en la elección de distribuidor. Los encuestados, ejecutivos de compras de empresas clientes de HATCO, puntúan a HATCO sobre cada atributo. La segunda clase de información hace referencia a los resultados de compras reales, bien sobre las evaluaciones de la satisfacción de los encuestados con HATCO, bien sobre el porcentaje de sus compras de productos a HATCO. La tercera clase de información contiene características generales de las empresas clientes (por ejemplo, tamaño de la empresa, tipo de industria).

Los datos proporcionados deberían dar a HATCO una mejor comprensión tanto de las características de sus clientes como de las relaciones entre sus percepciones de HATCO y sus acciones hacia HATCO (compras y satisfacción). En la Tabla 1.3 se proporciona una breve descripción de las variables de las bases de datos, que aparecen clasificadas como dependientes o independientes y métricas o no métricas. En el Apéndice A se proporciona un listado de las bases de datos para aquellos que deseen reproducir las soluciones a las que se llega en este libro. En las siguientes secciones se da una definición de cada variable y una explicación de su codificación.

TABLA 1.3. Descripción de las variables de la base de datos

Descripción de variable		*Tipo de variable*
PERCEPCIONES DE HATCO		
X_1	Velocidad de entrega	Métrica
X_2	Nivel de precios	Métrica
X_3	Flexibilidad de precios	Métrica
X_4	Imagen del fabricante	Métrica
X_5	Servicio conjunto	Métrica
X_6	Imagen de fuerza de ventas	Métrica
X_7	Calidad de producto	Métrica
RESULTADOS DE COMPRA		
X_9	Nivel de fidelidad	Métrica
X_{10}	Nivel de satisfacción	Métrica
CARACTERÍSTICAS DEL COMPRADOR		
X_8	Tamaño de empresa	No métrica
X_{11}	Compra al detalle	No métrica
X_{12}	Estructura de adquisición	No métrica
X_{13}	Tipo de industria	No métrica
X_{14}	Tipo de situación de compra	No métrica

Percepciones de HATCO

Cada una de estas variables se midió con una escala de puntuación gráfica, donde se dibujó una línea entre los puntos finales de diez centímetros, que se denominaron «Pobre» y «Excelente»:

Los encuestados indican sus ingresos haciendo una marca en la línea. La marca se midió y se registró la distancia desde cero (en centímetros). El resultado fue una escala que iba desde cero a diez, redondeado a un único decimal. Los siete atributos de HATCO puntuados por los encuestados son los siguientes:

- X_1 Velocidad de entrega — tiempo que transcurre hasta que se entrega el producto, una vez que se hubo confirmado el pedido.
- X_2 Nivel de precio — nivel de precios percibido por los clientes industriales.
- X_3 Flexibilidad de precios — la disposición percibida en los representantes de HATCO para negociar el precio de todas las compras.
- X_4 Imagen del fabricante — imagen conjunta del fabricante/distribuidor.
- X_5 Servicio — nivel conjunto de servicio necesario para mantener una relación satisfactoria entre el oferente y el comprador.
- X_6 Imagen de la fuerza de ventas — imagen conjunta de la fuerza de ventas del fabricante.
- X_7 Calidad del producto — nivel de calidad percibido en un producto particular (por ejemplo, el acabado o el rendimiento).

Resultados de compra

Se obtuvieron dos medidas específicas que reflejaron los resultados de las relaciones de compra del encuestado con HATCO. Estas medidas incluyen:

- X_9 Nivel de fidelidad — cuánto se compra a HATCO del total del producto de la empresa, medido en una escala de porcentaje de 100, que va desde 0 al 100 por cien.

X_{10} Nivel de satisfacción — satisfacción del comprador con las compras anteriores realizadas a HATCO, medidas en el mismo gráfico de la escala de clasificación de las entradas X_1 a X_7.

Características del comprador

Las cinco características utilizadas en el estudio, algunas métricas y otras no métricas, son las siguentes:

X_8 Tamaño de la empresa — tamaño de la empresa relativo respecto a otras empresas en el mismo mercado. Esta variable tiene dos categorías: 1 = grande y 0 = pequeña.

X_{11} Compra detallada — medida por la cual un comprador particular evalúa cada compra separadamente (análisis del valor total) o en función de una compra detallada, donde se especifican precisamente las características del producto deseado. Esta variable tiene dos categorías: 1 = emplea la aproximación al análisis del valor total, evaluando cada compra por separado y 0 = uso de la compra detallada.

X_{12} Estructura de la adquisición — método de adquisición/compra de productos a una compañía en particular. Esta variable tiene dos categorías: 1 = adquisición centralizada y 0 = adquisición descentralizada.

X_{13} Tipo de industria — clasificación de la industria a la que pertenece el comprador del producto. Esta variable tiene dos categorías: 1 = industria de la clase A y 0 = otras industrias.

X_{14} Tipo de situación de compra — tipo de situación a la que se enfrenta el comprador. Esta variable tiene tres categorías: 1 = nueva tarea, 2 = re-compra similar modificada y 3 = re-compra simple.

Otras bases de datos

En el texto se utilizan otras tres bases de datos especializadas. El Capítulo 2 emplea una base de datos más pequeña de varias de estas variables obtenidas en algunos estudios previamente realizados. El propósito es ilustrar la identificación de atípicos, manipulación de datos perdidos y comprobación de hipótesis estadísticas. Los Capítulos 8 y 10 examinan las bases de datos particularizadas para estas técnicas. En cada caso, la base de datos se describe más extensamente que en otros capítulos. La lista completa de estas bases de datos se da en el Apéndice A.

Organización de los capítulos restantes

Los capítulos restantes del texto se organizan en cuatro secciones, presentando cada uno de ellos una fase independiente en la ejecución del análisis multivariante.

- ***Sección 1: Preparación para el análisis multivariante*** se pronuncia sobre cuestiones que deben ser resueltas antes de la ejecución del análisis multivariante. Esta sección comienza con el Capítulo 2, que aborda los temas de decisiones sobre datos ausentes, cumplimiento de supuestos estadísticos subyacentes e identificación de factores que podrían tener una influencia desproporcionada sobre los resultados. El Capítulo 3 cubre el análisis factorial, una técnica que resulta especialmente adecuada para examinar las relaciones entre variables y oportunidades para crear escalas sumadas. Se combinan estos dos capítulos para proporcionar al investigador tanto las herramientas de diagnosis necesarias para la preparación de los datos, como las medidas para la reducción de datos y construcción de escalas que se puedan incluir en otras técnicas de análisis multivariante.

- ***Sección 2: Técnicas de dependencia*** aborda cinco técnicas de dependencia–regresión múltiple, análisis discriminante, análisis multivariante de la varianza, análisis conjunto y correlación canónica (Capítulos 4-8, respectivamente). Las técnicas de dependencia, tal y como se ha comentado antes, permiten al investigador valorar el grado de relación entre los variables dependientes e independientes. Las técnicas de dependencia varían según el tipo y naturaleza de la relación, tal y como se refleja en las propiedades de medición de las variables dependientes e independientes. Se examina cada técnica en función de la valoración de las relaciones de dependencia y de su capacidad de afrontar un tipo de objetivo de investigación específico.
- ***Sección 3: Técnicas de interdependencia*** (Capítulos 9 y 10) abordan las técnicas del análisis cluster y escala multidimensional. Estas técnicas proporcionan al investigador unas herramientas que resultan particularmente adecuadas para valorar la estructura, dado que se concentran en la descripción de las relaciones entre objetos, tanto sujetos (análisis cluster) como objetos tales como empresas, productos, etc., (escala multidimensional). Cabe notar que en la Sección 1 ya se ha tratado una de las técnicas de interdependencia principales, el análisis factorial, así como su capacidad de valorar la relación entre variables.
- ***Sección 4: Técnicas avanzadas y emergentes*** (Capítulos 11 y 12) proporcionan al investigador la introducción a un técnica multivariante avanzada de uso generalizado, los modelos de ecuaciones estructurales, junto con nuevas técnicas emergentes en las áreas de la búsqueda de datos, redes neuronales y «bootstrapping». Los objetivos de estos dos capítulos no es presentar un tratamiento completo de las cuestiones y usos de estas técnicas, sino proporcionar una introducción que permita al investigador el uso potencial de estas técnicas en situaciones de investigación específicas.

Resumen

Este capítulo ha introducido el tema del análisis de datos multivariante. Los capítulos siguientes describen cada una de las técnicas en suficiente detalle como para permitir al analista novel comprender qué es lo que puede conseguir una técnica particular, cuando y cómo deberían aplicarse y cómo interpretar los resultados de su interpretación. Las lecturas situadas al final de cada capítulo, extraídas de la literatura académica española, mostrarán la aplicación y la interpretación de las técnicas.

Cuestiones de revisión

1. Defina el análisis multivariante con sus propias palabras.
2. Nombre varios factores que han contribuido a aumentar la aplicación de las técnicas del análisis de datos multivariantes en los últimos años.
3. Enumere y describa las técnicas de análisis de datos multivariante descritas en este capítulo. Cite ejemplos adecuados para cada técnica.
4. Explique cómo y por qué los diversos métodos multivariantes pueden verse como una familia de técnicas.
5. ¿Por qué es importante el conocimiento de las escalas de medida para entender el análisis de datos multivariante?
6. ¿Cuáles son las diferencias entre la significación práctica y estadística? ¿Es un prerrequisito de la otra?
7. ¿Cuáles son las implicaciones de una potencia estadística baja? ¿Cómo puede aumentarse la potencia si se considera muy baja?
8. Explique detalladamente la aproximación al diseño de modelos multivariantes, centrándose en los supuestos principales en cada paso.

Bibliografía

1. Bearden, William O., Richard G. Netemeyer, and Mary F. Mobley (1993), *Handbook of Marketing Scales, Multi-Item Measures for Marketing and Consumer Behavior.* Newbury Park, Calif.: Sage.
2. BMDP Statistical Software, Inc. (1991), *SOLO Power Analysis.* Los Angeles.
3. Brent, Edward E., Edward J. Mirielli, and Alan Thompson (1993), *Ex-Sample™: An Expert System to Assist in Determining Sample Size, Version 3.0.* Columbia, Mo.: Idea Works.
4. Brent, Edward E., et al. (1991), *Statistical Navigator Professional™: An Expert System to Assist in Selecting Appropriate Statistical Analyses, Version 1.0.* Columbia, Mo.: Idea Works.
5. Brunner, Gordon C., and Paul J. Hensel (1993), *Marketing Scales Handbook, A Compilation of Multi-Item Measures*. Chicago: American Marketing Association.
6. Cohen, J. (1977), *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.
7. Gatty, R. (1966), «Multivariate Analysis for Marketing Research: An Evaluation.» *Applied Statistics* 15 (November): 157–172.
8. Hardyck, C. D., and L. F. Petrinovich (1976), *Introduction to Statistics for the Behavioral Sciences*, 2d ed. Philadelphia: Saunders.
9. Mooney, Christopher Z., and Robert D. Duval (1993), *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Beverly Hills, Calif.: Sage.
10. Sullivan, John L., and Stanley Feldman (1979), *Multiple Indicators: An Introduction.* Beverly Hills, Calif.: Sage.

SECCIÓN

1

PREPARACIÓN PARA EL ANÁLISIS MULTIVARIANTE

Valoración general

La Sección 1 proporciona una serie de herramientas y análisis que ayudan a preparar al investigador para la naturaleza crecientemente compleja del análisis multivariante. El investigador prudente aprecia la necesidad de un conocimiento más profundo de los datos, tanto en términos estadísticos como de concepto. Aunque las técnicas multivariantes que se abordan en este texto ofrecen al investigador una serie de herramientas poderosas de análisis, también presentan el riesgo de (a) separar el investigador aún más de la posibilidad de adquirir una comprensión sólida de los datos y (b) conducir a nociones erróneas que presentan una manera «fácil y rápida» de identificar relaciones. Dado que el investigador depende más de estas técnicas para encontrar la respuesta y menos de una base conceptual y de comprensión de las propiedades fundamentales de los datos, aumenta el riesgo de que aparezcan problemas serios de aplicación errónea de las técnicas, de violación de las propiedades estadísticas, o de inferencia e interpretación de los resultados erróneos. Nunca se pueden eliminar completamente estos riesgos. No obstante, las herramientas y análisis abordados en esta sección mejorarán la capacidad del investigador a la hora de reconocer muchos de estos problemas conforme aparecen y aplicar el remedio apropiado.

Capítulos de la Sección 1

Esta sección comienza con el Capítulo 2, Análisis previo de los datos, que aborda los temas de decisión sobre datos ausentes, cumplimiento de los supuestos estadísticos subyacentes e identificación de factores que podrían tener un impacto desproporcionado sobre los resultados. Estos análisis proporcionan valoraciones empíricas simples que detallan las propiedades estadísticas críticas de los datos. El Capítulo 3, referido al análisis factorial, presenta un debate sobre una técnica de interdependencia particularmente adecuada para examinar las relaciones entre las variables y la creación de escalas sumadas. La «búsqueda de la estructura» con el

análisis factorial puede revelar interrelaciones importantes entre variables y puede proporcionar una base objetiva para el desarrollo del modelo conceptual. Por tanto, los dos capítulos de esta sección se combinan para proporcionar al investigador no solamente las herramientas necesarias para preparar los datos para el análisis, sino también la manera apropiada de reducción de datos y construcción de escalas que pueden mejorar notablemente otras técnicas de análisis multivariante.

CAPÍTULO 2

Análisis previo de los datos

OBJETIVOS DE APRENDIZAJE

- Seleccionar el método gráfico apropiado para examinar las características de los datos o relaciones de interés.
- Comprender los diferentes tipos de procesos de datos ausentes.
- Evaluar el tipo y el impacto potencial de los datos ausentes.
- Explicar las ventajas y desventajas de los enfoques disponibles para tratar con los datos ausentes.
- Identificar los casos atípicos univariantes, bivariantes y multivariantes.
- Comprobar en los datos la mayor parte de los supuestos de las técnicas multivariantes.
- Determinar el mejor método de transformación de los datos dado un problema específico.
- Comprender la manera de incorporar variables no métricas como datos métricos.

En este capítulo revisaremos y describiremos los métodos habitualmente utilizados para examinar los datos. El examen de los datos es un paso necesario, que lleva tiempo, y que habitualmente se descuida por parte de los analistas de datos. Un análisis cuidadoso de los datos conduce a una mejor predicción y a una evaluación más precisa de la dimensionalidad. La sección introductoria de este capítulo ofrece un resumen de varias técnicas gráficas que se ponen a disposición del investigador como medio para presentar los datos. Estas técnicas proporcionan al analista un conjunto de formas simple, aunque completo, de examinar tanto las variables individuales como las relaciones entre ellas. Otros problemas que conciernen al analista cuando está examinando los datos son cómo evaluar y solucionar los problemas en el diseño de la investigación y en la recogida de datos. Más concretamente, este capítulo se dirige a la evaluación de los datos ausentes, identificación de casos atípicos y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes. Los datos ausentes son una molestia para los investigadores. Pueden ser producto

de errores en la introducción de los datos o de la omisión de respuestas por parte de los encuestados. En este capítulo se discutirá la clasificación de los datos ausentes y los procesos o razones que explican su presencia. Los casos atípicos, o respuestas extremas, pueden influenciar indebidamente el resultado de un análisis multivariante. Por esta razón, se discuten a continuación los métodos para evaluar su impacto. Finalmente, se revisan los supuestos que subyacen en la mayor parte de los análisis multivariantes. Para aplicar cualquier técnica multivariante, el investigador debe evaluar el ajuste de la muestra de datos con los supuestos subyacentes en la técnica multivariante. Por ejemplo, los investigadores que desean analizar el análisis de regresión (Capítulo 4) estarán particularmente interesados en evaluar los supuestos de normalidad, homocedasticidad, independencia del error y linealidad. Cada una de estas cuestiones deberían ser abordadas en cierta medida para cada aplicación de la técnica multivariante.

Adicionalmente, en este capítulo se introduce al investigador en los métodos para incorporar las variables no métricas en aplicaciones que requieren variables métricas mediante la creación de un tipo de variable métrica especial conocida como variable ficticia. La aplicabilidad del uso de las variables ficticias varía con cada proyecto de análisis de datos.

Introducción

Las tareas implícitas en el examen previo de los datos pueden parecer insignificantes y sin consecuencias a primera vista, pero son una parte esencial del análisis multivariante. Si bien las técnicas multivariantes suponen un tremendo poder analítico en manos del investigador, también crean una gran carga para éste, ya que tiene que asegurarse de que se mantengan los cimientos teóricos y estadísticos sobre las que se basan. Examinando con anterioridad los datos de la aplicación de una técnica multivariante, el investigador gana una comprensión básica de los datos y las relaciones entre las variables. En primer lugar, el investigador obtiene un conocimiento básico de los datos y las relaciones entre las variables. Las técnicas multivariantes plantean grandes demandas al analista en la compresión, interpretación y articulación de resultados basados en relaciones cuya complejidad está en continuo aumento. El conocimiento de las interrelaciones de variables puede ayudar enormemente en la especificación y refinamiento del modelo multivariante, así como proporcionar una perspectiva razonable para la interpretación de los resultados. En segundo lugar, las técnicas multivariantes demandan mucho más de los datos que se van a analizar. La potencia estadística de las técnicas multivariantes requiere grandes conjuntos de datos y supuestos más complejos que los que encontramos en los análisis univariantes. La sofisticación analítica necesaria para asegurar que se consiguen los requerimientos estadísticos al aplicar la técnica multivariante elegida, obliga al analista a usar una serie de técnicas de examen de los datos que en muchas ocasiones rivaliza en complejidad con la propia técnica multivariante. Además, los efectos de los datos ausentes, los cuales por definición no se representan directamente en los resultados, pueden ser sustanciales por el impacto que tienen sobre la naturaleza y carácter de los resultados. El propósito de este capítulo es proporcionar una visión general de las técnicas de examen de los datos, que van desde el simple proceso de inspección visual de los gráficos al proceso estadístico multivariante que requiere el análisis de datos ausentes y a la comprobación de los supuestos subyacentes en todos los métodos multivariantes.

Tanto los analistas noveles como los experimentados pueden sentirse tentados de leer por encima o incluso saltarse este capítulo, para ir directamente a los conocimientos referentes a la aplicación de una técnica o técnicas multivariantes. Aunque pueda parecer que es malgastar el tiempo, el esfuerzo y los recursos dedicados al proceso de examen de los datos, el investigador debe ver estas técnicas como una «inversión en un seguro multivariante». Incluso aunque una técnica sirva para hacer una estimación adecuada y obtener resultados, los problemas «ocultos» que surgen de las cuestiones expuestas en este capítulo pueden conducir a problemas potencialmente catastróficos. Se pueden evitar estos problemas al seguir estos análisis cada vez que se aplique una técnica multivariante. Estos esfuerzos recompensarán el esfuerzo a largo plazo, evitando cometer

un error serio y posiblemente fatal. Desde aquí animamos al lector a adoptar estas técnicas antes de verse obligado a rectificar posteriormente a causa de los problemas surgidos durante el análisis.

El capítulo nos remite a cuatro fases distintas del examen de los datos. Éstas incluyen: (1) un examen gráfico de la naturaleza de las variables a analizar y las relaciones que forman las bases del análisis multivariante, (2) un proceso de evaluación para entender el impacto que pueden tener los datos ausentes sobre el análisis, y una serie de alternativas para casos reiterados de ausencia de datos en el análisis, (3) las técnicas que mejor se ajustan para la identificación de casos atípicos, aquellos casos que por su singularidad pueden distorsionar las relaciones sobre una o más variables estudiadas y (4) los métodos analíticos necesarios para evaluar adecuadamente la capacidad de los datos para cumplir los supuestos estadísticos específicos de muchas técnicas multivariantes. Este capítulo se concluye con la introducción de una técnica para incorporar variables no métricas cuando se requieren variables métricas. Se crean una serie de variables métricas de reemplazo para representar las categorías de las variables no métricas.

Examen gráfico de los datos

Como se ha discutido antes, el uso de las técnicas multivariantes supone una mayor carga para el investigador a la hora de entender, evaluar e interpretar los resultados más complejos. En estas tareas, la ayuda proviene de la comprensión de las características básicas de los datos y de sus relaciones subyacentes. Cuando se considera un análisis univariante, el nivel de comprensión es muy simple. Pero a medida que el investigador se introduce en los más complejos análisis multivariantes, la necesidad y el nivel de comprensión aumentan considerablemente. Esta sección revisa algunos de los métodos gráficos básicos que existen para ayudar a la obtención de una comprensión básica de las características de los datos, particularmente en un sentido multivariante.

La aparición y la difusión del uso de programas estadísticos diseñados para ordenadores personales ha llevado a una mejora del acceso a tales métodos. La mayoría de los programas estadísticos tienen módulos particularmente comprensivos de técnicas gráficas preparadas para el examen de los datos que se ven mejoradas muchas veces con medidas estadísticas más detalladas de la descripción de los datos. Las siguientes secciones ilustrarán algunas de las técnicas usadas de forma más común en el examen de las características de la distribución, relaciones bivariantes, diferencias por grupos e incluso perfiles multivariantes.

La naturaleza de la variable: Análisis de la forma de la distribución

El punto de partida para entender la naturaleza de cualquier variable es caracterizar la forma de su distribución. Algunas medidas estadísticas se discutirán en una sección posterior dedicada a la normalidad. No obstante, muchas veces el investigador puede obtener una perspectiva adecuada de la variable a través de un **histograma**. Un histograma es una representación gráfica de los datos que muestra la frecuencia de los casos (valores de los datos) en categorías de datos. Las frecuencias se representan para examinar la forma de la distribución de respuestas. Si el rango de respuestas va de 1 a 10, el investigador puede construir un histograma contando el número de respuestas que fueron un uno, un dos, etc. Para las variables continuas se forman categorías, dentro de las cuales la frecuencia de los valores de datos está tabulada. Por ejemplo, las respuestas para X_1 de la base de datos introducida en el Capítulo 1 se muestran en la Figura 2.1. Se usan las categorías de puntos intermedios 0,0, 0,5, 1,0, 1,5..., hasta 6,0. La altura de las barras representa la frecuencia de los valores de los datos en cada categoría. Si el examen de la distribución tiene como objetivo evaluar su normalidad (para más detalles sobre este asunto véase una sección posterior sobre los supuestos de comprobación), se puede superponer la curva normal sobre la distribución, como se ha he-

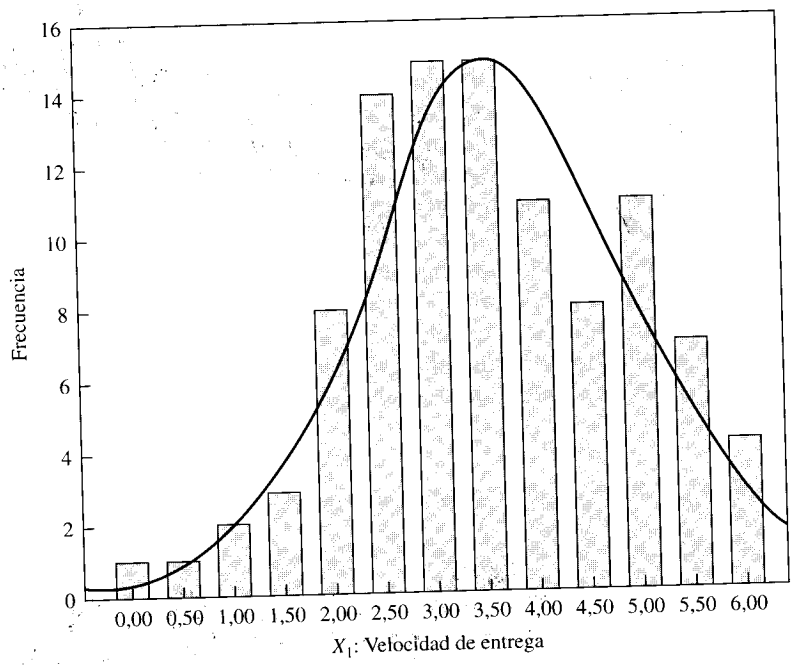


FIGURA 2.1. Una representación gráfica de una distribución univariante: El histograma.

cho en la Figura 2.1. El histograma puede utilizarse para examinar cualquier tipo de variable, desde los valores originales a los residuos de una técnica multivariante.

Una variante del histograma es **el diagrama de tallo y hojas** que presenta el mismo cuadro gráfico pero que también proporciona una enumeración de los valores de los datos reales. El diagrama de tallo y hojas de la Figura 2.2 se compone de tallo y hojas. El tallo es el valor raíz, al cual se le añaden hojas. Por ejemplo, en la Figura 2.2, el primer tallo es 0,0. A esto se le añade la hoja de 0, resultando en un valor de cero.

En el siguiente tallo, a un valor de 0,6 se le añade el tallo de cero, resultante en un valor de 0,6. Si se compilan las frecuencias de X_1, 0,0 y 0,6 son los dos primeros valores. En el otro extremo de la figura, el tallo es 6,0. Se asocia con dos hojas (0 y 1), representando los valores 6,0 y 6,1. És-

Frecuencia	Tallo y hojas		
1,00	0	*	0
1,00	0	.	6
3,00	1	*	013
7,00	1	.	6688999
12,00	2	*	001333444444
10,00	2	.	5566788899
18,00	3	*	000001111233444444
10,00	3	.	5666777889
10,00	4	*	001122233
10,00	4	.	556778999
11,00	5	*	00112223344
5,00	5	.	55689
2,00	6	*	01

Anchura del tallo: 1,0
Cada hoja: 1 caso (s)

Casos válidos: 100,0 Casos perdidos: 0,0 Porcentaje perdido: 0,0

FIGURA 2.2. Gráfico de tallo y hojas de X_1 (velocidad de entrega).

tos son los dos mayores valores para X_1. El diagrama de hojas y tallos proporciona un perfil general de la distribución de la misma forma que se encuentra en el histograma, además de los valores de los datos reales.

Análisis de relación entre variables

El examen de la distribución de una variable es esencial, pero muchas veces el investigador está interesado en examinar las relaciones entre dos o más variables. El método más popular de análisis de las relaciones bivariantes es el **gráfico de dispersión**, un gráfico de puntos de datos basados en dos variables. Se presenta una variable en el eje horizontal y la otra en el vertical. Las variables pueden ser valores observados, valores esperados o incluso **residuos**. Los puntos del gráfico representan los correspondientes valores conjuntos de las variables para cualquier caso dado. El patrón de los puntos representa la relación entre las variables. Cuando los puntos se organizan a lo largo de una línea recta, tenemos una relación lineal de correlación. Un conjunto de puntos curvados puede indicar una relación no lineal, que se puede tratar de varias formas (véase la ulterior discusión sobre linealidad). O puede que no existan patrones, sólo un conjunto de puntos aparentemente aleatorios. En este caso, no hay relación.

Existen muchos tipos de gráficos de dispersión, pero un formato que se ajusta particularmente a las técnicas multivariantes es la matriz del gráfico de dispersión (véase Figura 2.3). Aquí se representa el gráfico de dispersión para todas las combinaciones de variables en la porción inferior de la matriz. Las diagonales contienen los histogramas de las variables. En la parte superior de la matriz se incluyen las correlaciones correspondientes para que el lector pueda valorar la correlación representada en cada gráfico de dispersión. La Figura 2.3 presenta los gráficos de dispersión de un conjunto de variables de la base de datos de HATCO (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 y X_9). Por ejemplo, el gráfico de dispersión entre X_1 y X_9 de la esquina de abajo a la izquierda representa una correlación de 0,676. Los puntos están muy alineados alrededor de una línea recta, indicando una elevada correlación. El gráfico de dispersión de la columna más a la izquierda y tercero desde arriba (X_1 versus X_4) demuestra lo contrario, una ausencia casi total de relación como evidencia la amplia dispersión de los puntos y la correlación de 0,050. Las matrices de gráficos de dispersión y los gráficos individuales se encuentran ya en todos los programas de estadística. Discutiremos en una sección posterior una variante del gráfico de dispersión para detectar casos atípicos, donde sobrepondremos una elipse que representa un intervalo de confianza para una distribución normal bivariante para poder identificar los casos atípicos.

Análisis de las diferencias entre grupos

El investigador se enfrenta también con la tarea de entender el carácter y la diferencia entre dos o más grupos de una variable para dos o más variables métricas, tal y como ocurre en el análisis discriminante, análisis de la varianza y análisis multivariante de la varianza. En estos casos, el investigador necesita entender cómo se distribuyen los valores para cada grupo y si existen suficientes diferencias entre ellos como para tener significación estadística. Otro aspecto importante es identificar los **casos atípicos** que pueden resultar ser aparentes sólo cuando los valores de los datos se separan en grupos. El método que se utiliza para esta tarea es el gráfico de cajas **(boxplot)**, una representación gráfica de la distribución de los datos. Los límites superior e inferior de la caja marcan los cuartiles superior e inferior de la distribución de los datos. Por tanto, la longitud de la caja es la distancia entre el primer y el tercer cuartil, de forma que la caja contiene el 50 por ciento de los datos centrales de la distribución. La línea dentro de la caja señala la posición de la mediana. Si ésta cae cerca del final de la caja, se indica la presencia de asimetría. Cuanto mayor es la caja, mayor es la extensión de las observaciones. Las líneas que se extienden desde cada caja (llamadas *bigotes*) representan la distancia entre la mayor y la menor de las observaciones que están a menos de un cuartil de la caja. Estos valores están marcados con una X. Los casos atípicos son observaciones que se sitúan entre 1,0 y 1,5 cuartiles fuera de la caja.

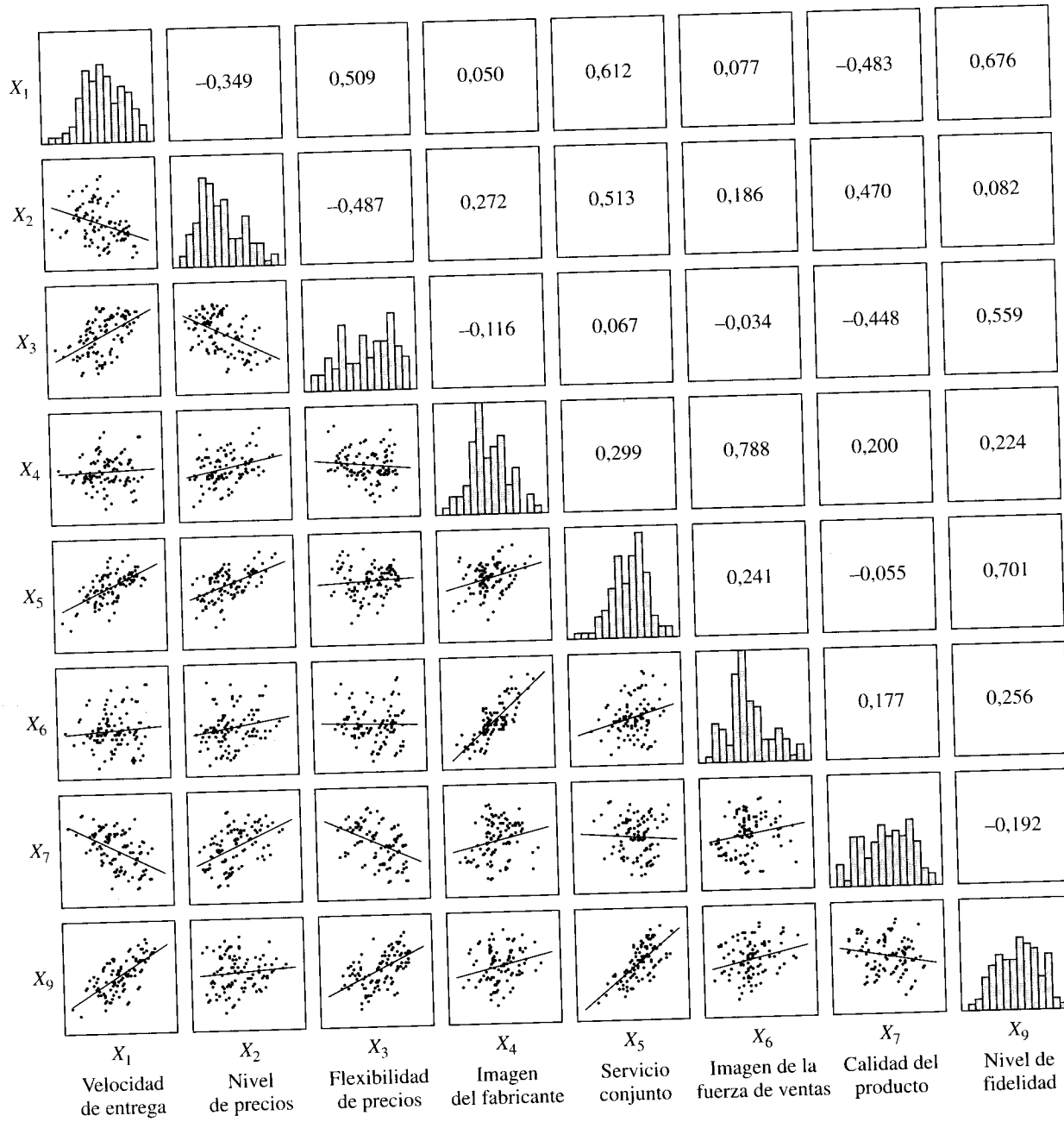


FIGURA 2.3. Matriz del gráfico de dispersión de variables métricas.
Nota: Los valores por encima de la diagonal son correlaciones bivariantes, con gráficos de dispersión correspondientes por debajo de la diagonal. La diagonal presenta la distribución de cada variable.

Los valores extremos son aquellas observaciones mayores que están a 1,5 cuartiles fuera de los límites de la caja.

La Figura 2.4 muestra el gráfico de cajas para X_1 (velocidad de entrega) para cada grupo de X_{14} (tipos de situaciones de compra) de la base de datos de HATCO. Los tres grupos tienen un conjunto de valores muy diferente, lo que indica que existen verdaderas diferencias entre los grupos en términos de percepciones de la velocidad de entrega. El gráfico para el primer tipo de situación de compra también indica que existe un caso atípico. El investigador debería examinar esta observación y considerar las posibles soluciones. Las mismas para la existencia de casos atípicos se discutirán en profundidad posteriormente.

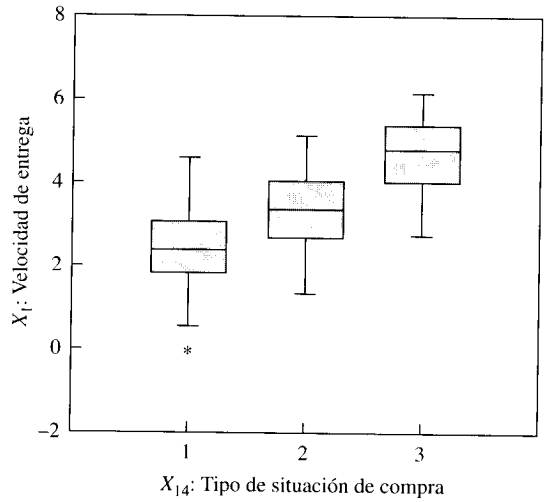


FIGURA 2.4. Gráfico de cajas y bigotes.

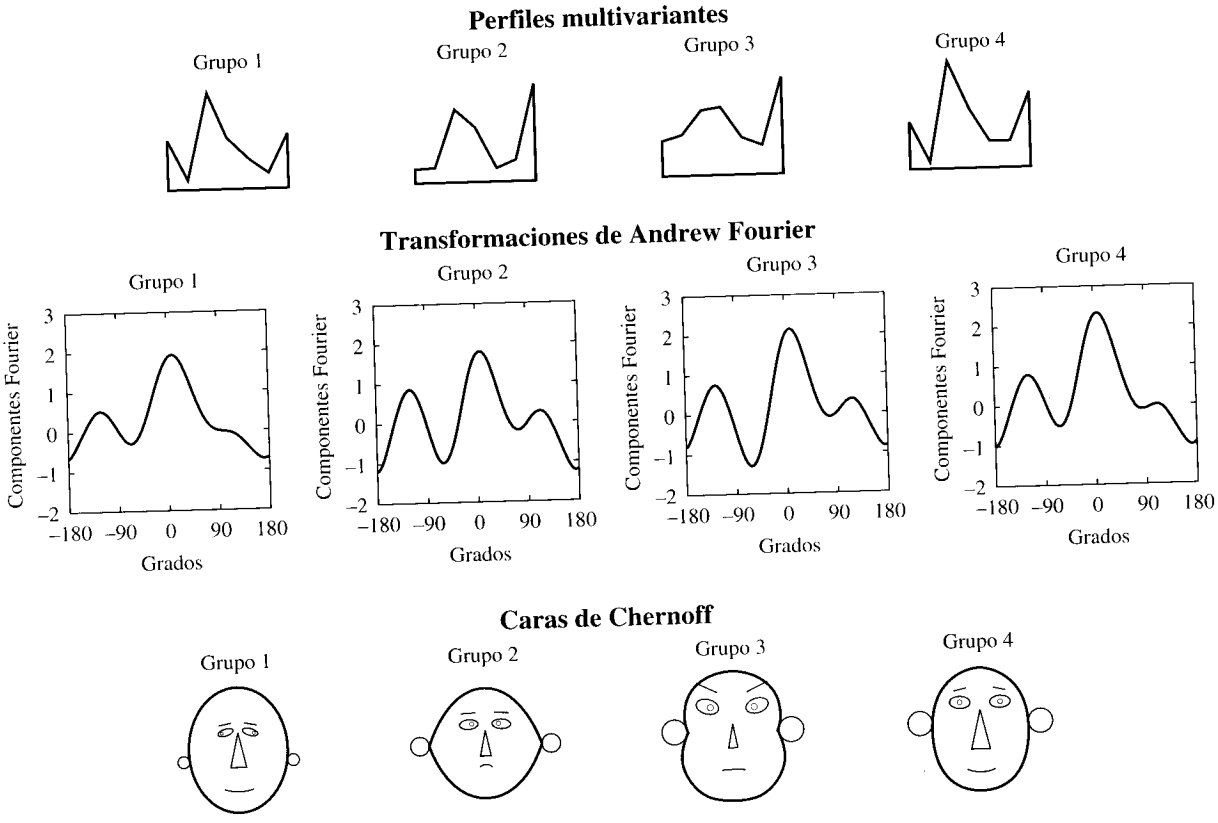
Perfiles multivariantes

Hasta ahora, los métodos gráficos se han ceñido a representaciones univariantes o bivariantes. Pero en muchas instancias, el investigador puede desear comparar observaciones caracterizadas por la presencia de más de dos variables, necesitando un medio de presentar un perfil multivariante de una observación, tanto si es para propósitos descriptivos como si es un complemento para procedimientos analíticos. Para salvar estas dificultades se han elaborado varios métodos de gráficos multivariantes que se centran en una de estas tres aproximaciones [7]. Los primeros son (a) los glyphs o metroglyphs, que son un tipo de círculo con radios que corresponden a los valores de los datos o (b) un perfil multivariante, que representa un diagrama de barras para cada observación. Una segunda forma de representación multivariante implica una transformación matemática de los datos originales en una relación matemática, que pueda ser representada gráficamente. La técnica más común es la transformación de Andrew Fourier. La última aproximación es el uso de representaciones gráficas con representatividad icónica, siendo la más popular la cara [3]. El valor de este tipo de representación es la inherente capacidad que tienen los seres humanos para procesar su interpretación. Como señala Chernoff [3, pag. 9]:

> Creo que aprendemos muy pronto a estudiar y reaccionar a los rostros. Nuestro catálogo de respuestas a los rostros agota una gran parte de nuestro diccionario de emociones e ideas. Percibimos las caras como un espíritu y nuestro «ordenador» es muy rápido en obtener la información relevante y filtrar el ruido cuando nos enfrentamos a un número limitado de caras.

Las representaciones faciales proporcionan un potente formato gráfico pero también dan lugar a un número de consideraciones entre las que se encuentran la asignación de variables a caracteres faciales, percepciones involuntarias y la cantidad de información que realmente puede ser tratada. La discusión de estos asuntos está más allá del objetivo de este texto y aconsejamos a los lectores interesados que los repasen antes de utilizar estos métodos [10, 11].

La Figura 2.5 contiene una ilustración de estos tres tipos de **representaciones gráficas multivariantes** que fueron realizadas con SYSTAT, pero que se pueden obtener con muchos programas estadísticos de ordenador personal. La parte superior de la Figura 2.5 contiene ejemplos de cada una de las representaciones gráficas multivariantes: perfiles, transformaciones de Fourier y caras icónicas. Los valores de cuatro observaciones sobre siete variables están en una tabla al final de la figura. En este ejemplo, los datos son perfiles de los cuatro grupos de clientes respecto de los siete factores de resultados de la base de datos de HATCO. A partir de los valores de los datos



Datos reales

Grupo	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	4,794	1,622	8,267	4,717	3,222	2,067	5,044
2	2,011	2,133	6,544	5,267	2,039	2,672	8,483
3	3,700	4,158	6,008	6,242	3,900	3,233	8,258
4	4,809	1,510	9,319	5,690	3,148	3,195	6,981

FIGURA 2.5. Ejemplos de representación gráfica multivariante.

reales, las similitudes y las diferencias son difíciles de distinguir, incluso en la medida en que puede que no existan diferencias. El objetivo de los perfiles multivariantes es representar los datos de tal forma que permitan identificar fácilmente las similitudes y las diferencias. La primera representación consiste en perfiles multivariantes, que muestran que la parte situada en el extremo izquierdo de la distribución es muy baja para el grupo dos, y alta para los grupos uno y cuatro. Este patrón corresponde a los valores de X_1, y se pueden hacer comparaciones entre los grupos en una sola variable o a través de variables para un solo grupo. El segundo tipo de representación gráfica multivariante en la figura es la transformación de Andrew Fourier, que representa los valores mediante una expresión matemática. Aunque las comparaciones para un valor único son más difíciles, esta forma de representación gráfica nos presenta en un solo gráfico una comparación generalizada y la agrupación de observaciones. Finalmente, se han construido los símbolos icónicos (las caras de Chernoff) con las siete variables asignadas a las características faciales. En este ejemplo, X_1 controla la representación de la boca, se asigna X_2 para las facciones de cejas, se asigna X_3 para las características de la nariz, se asigna X_4 para los ojos, X_5 controla la forma de la cara, X_6 se refiere a la oreja y se asigna X_7 a la posición de la pupila. Con relación a X_1, los grupos uno

y cuatro tienen sonrisas y el grupo dos tiene un ceño. Esto corresponde a valores grandes para los grupos uno y cuatro y valores pequeños para el grupo dos. Esta forma de representación gráfica combina la capacidad de hacer comparaciones específicas entre grupos o dentro de los grupos vistos en los perfiles además de las comparaciones globales más generalizadas que se encuentran en las transformaciones de Andrew Fourier. El investigador puede emplear cualquiera de estos métodos cuando examina datos multivariantes para proporcionar un formato que es muchas veces más penetrante que una simple revisión de los datos reales.

Resumen

No hay que entender las representaciones gráficas de esta sección como un sustituto de las medidas de diagnóstico estadístico discutidas en las últimas secciones de este capítulo y en otros capítulos. Pero proporcionan una forma alternativa de desarrollar una perspectiva del carácter de los datos y de las interrelaciones que existen, incluso si son multivariantes en su naturaleza. El antiguo refrán «una imagen vale más que mil palabras» es verdad muchas veces en el uso de las representaciones gráficas para las aplicaciones comparativas o de diagnóstico.

Datos ausentes

Los **datos ausentes** son algo habitual en el análisis multivariante; de hecho, raramente el investigador evita enfrentarse con alguna forma del problema de los datos ausentes. Por esta razón, el desafío consiste en enfrentarse a los resultados producidos por los datos ausentes en los procesos de estimación y que afectan a la generalidad de los resultados. Para hacer esto, la ocupación primaria del investigador es determinar las razones que subyacen en el dato ausente, siendo en muchos casos la extensión de la ausencia de datos una cuestión secundaria. Esta necesidad de centrarse en las razones de la ausencia de datos se desprende del hecho de que el investigador debe entender el proceso principal de esta ausencia de datos para seleccionar el curso de acción apropiado.

Un **proceso de datos** ausentes es cualquier evento sistemático externo al encuestado (tales como errores en la introducción de los datos o problemas de su recolección) o acción por parte del encuestado (tales como rehusar a contestar) que da lugar a la ausencia de datos. Los efectos de ciertos procesos de ausencia de datos son conocidos y directamente introducidos en el plan de investigación. Pero hay otros, particularmente aquellos basados en acciones por parte del encuestado, que rara vez son conocidos. Cuando los procesos de ausencia de datos son desconocidos, el investigador intenta identificar cualquier patrón en los datos ausentes que caracterizarían dicho proceso. Al hacerlo, plantea cuestiones tales como: (1) ¿están los datos ausentes distribuidos aleatoriamente entre las observaciones o se pueden identificar distintas pautas? y (2) ¿en qué medida son relevantes? Si se encuentran pautas y la extensión de los datos ausentes es suficiente como para garantizar un curso de acción, entonces se asume que está operando algún proceso de ausencia de datos y que algunos de los resultados estadísticos basados en esos datos podrían estar sesgados en la medida en que las variables incluidas en el análisis están influidas por los procesos de pérdida de datos. El interés por entender los procesos de ausencia de datos es similar a la necesidad de entender las causas de la ausencia de respuestas en el proceso de recolección de datos. Por ejemplo, ¿son diferentes aquellas personas que no respondieron de las personas que sí lo hicieron? Si es así, ¿tienen algún impacto estas diferencias en el análisis, los resultados o su interpretación? Asuntos similares a éstos surgen también de la inexistencia de respuestas a variables individuales.

El impacto de los datos ausentes es perjudicial no sólo por sus potenciales sesgos «escondidos» sino también por su efecto en el tamaño de la muestra disponible para el análisis. Por ejemplo, si no se aplican soluciones para la ausencia de datos, ninguna observación con datos ausentes sobre cualquiera de las variables será incluida en el análisis. En muchos análisis multivariantes, en par-

ticular en aplicaciones a investigaciones genéricas, la ausencia de datos elimina tantas observaciones que lo que en principio era una muestra adecuada se reduce a una inadecuada. En tales situaciones, el investigador debe buscar observaciones adicionales o encontrar una solución para la ausencia de datos en la muestra original. Aunque encontrar un remedio para la ausencia de datos es la solución más práctica, existen pocas guías para el diagnóstico y solución de la ausencia de datos. Por esta razón, las siguientes secciones discuten los diferentes tipos de procesos de ausencia de datos, métodos para identificar su naturaleza y las soluciones existentes para dar cabida a la ausencia de datos en el análisis multivariante.

Un ejemplo sencillo del análisis de datos ausentes

La Tabla 2.1 contiene un ejemplo sencillo de datos ausentes entre 20 casos. De la misma forma que resulta típico para muchas series de datos, especialmente en la investigación de encuestas, el número de datos ausentes varía mucho entre los casos y las variables. En este ejemplo, podemos ver que todas las variables (V_1 a V_5) tienen algunos datos ausentes, en V_3 falta más de la mitad (55 por ciento) de todos los valores. Tres casos (3, 13 y 15) tienen más del 50 por ciento de datos ausentes y sólo cinco casos tienen datos completos. En conjunto, un 22 por ciento de los valores están ausentes. Si se hiciera un análisis multivariante que necesitara datos completos, los datos se verían reducidos a solamente cinco casos, demasiado poco para cualquier tipo de análisis. Este nivel de reducción en los casos disponibles es frecuente en muchas aplicaciones.

Las soluciones más sofisticadas para datos ausentes se abordarán con detalle en secciones posteriores. No obstante, una opción obvia es la eliminación de las variables y/o casos. En nuestro ejemplo, si suponemos que los fundamentos conceptuales de la investigación no se alteran sustancialmente por la supresión de una variable, la eliminación de V_3 es una manera de reducir el número de datos ausentes. Solamente con la supresión de V_3, siete casos adicionales, de un total de 12, ahora tienen

TABLA 2.1. Ejemplo hipotético de datos ausentes

						Datos ausentes por caso	
Caso ID	V_1	V_2	V_3	V_4	V_5	*Número*	*Porcentaje*
1	1,3	9,9	6,7	3,0	2,6	0	0
2	4,1	5,7			2,9	2	40
3		9,9		3,0		3	60
4	0,9	8,6		2,1	1,8	1	20
5	0,4	8,3		1,2	1,7	1	20
6	1,5	6,7	4,8		2,5	1	20
7	0,2	8,8	4,5	3,0	2,4	0	0
8	2,1	8,0	3,0	3,8	1,4	0	0
9	1,8	7,6		3,2	2,5	1	20
10	4,5	8,0		3,3	2,2	1	20
11	2,5	9,2		3,3	3,9	1	20
12	4,5	6,4	5,3	3,0	2,5	0	9
13					2,7	4	80
14	2,8	6,1	6,4		3,8	1	20
15	3,7			3,0		3	60
16	1,6	6,4	5,0		2,1	1	20
17	0,5	9,2		3,3	2,8	1	20
18	2,8	5,2	5,0		2,7	1	20
19	2,2	6,7		2,6	2,9	1	20
20	1,8	9,0	5,0	2,2	3,0	0	0
DATOS AUSENTES POR VARIABLE						TOTAL VALORES AUSENTES	
Número	2	2	11	6	2	Número: 23	
Porcentaje	10	10	55	30	10	Porcentaje: 23	

información completa. Si se eliminan los tres casos (3, 13 y 15) con cantidades de datos ausentes excepcionalmente altas, el número total de datos ausentes se reduce ahora a solamente cinco casos, un 7,4 por ciento de todos los valores. No obstante, todos estos cinco datos ausentes están presentes en V_4, y tenemos que buscar también cualquier patrón entre estos datos. Si comparamos los casos con los datos ausentes para V_4 con los que tienen valores V_4 válidos, podemos ver evolucionar un patrón con respecto al V_2. Los cinco casos con los valores ausentes para V_4 también tienen los valores más bajos para V_2, lo cual indica que los datos ausentes para V_4 están estrechamente asociados con los niveles más bajos de V_2. Esta asociación sistemática entre los datos ausentes y los datos válidos tiene un impacto directo sobre cualquier análisis en los que se incluyen V_4 y V_2. En este caso, el investigador siempre tiene que examinar los resultados que incluyen V_4 o V_2 para ver el posible impacto de este proceso de datos ausentes sobre los resultados.

Conocimiento de las razones causantes de datos ausentes

Antes de que se pueda instrumentar cualquier solución para la ausencia de datos, en primer lugar el investigador debe diagnosticar y comprender los procesos de ausencia de datos que subyacen en este fenómeno. Algunas veces estos procesos se encuentran bajo el control del investigador y pueden ser identificados explícitamente. En tales casos, la ausencia de casos se denomina **prescindible**, lo que significa que no se necesitan soluciones específicas para la ausencia de datos dado que los límites de la ausencia de los datos son inherentes a la técnica usada [9].

Datos ausentes prescindibles

Un ejemplo de proceso de datos ausentes prescindibles es el «dato ausente» de aquellas observaciones de una población que no están incluidas en la muestra. El propósito de la técnica multivariante es generalizar las observaciones de la muestra al conjunto de la población, lo que supone un intento por salvar los datos ausentes de las observaciones que no están en la muestra. El investigador hace prescindibles estos datos ausentes usando una muestra probabilística de los encuestados seleccionados. La muestra probabilística permite al investigador especificar que los procesos de datos ausentes causantes de las observaciones omitidas son aleatorios y que los datos ausentes pueden explicarse como un error muestral en los procedimientos estadísticos. Por tanto, los «datos ausentes» de las observaciones no seleccionadas son prescindibles.

Otro caso de datos ausentes prescindibles tiene lugar cuando los datos están censurados. Los **datos censurados** son observaciones incompletas como consecuencia de su etapa en el proceso de ausencia de datos. Un ejemplo típico es un análisis de las causas de fallecimiento. Los encuestados que todavía viven no pueden proporcionar información completa (es decir, causa de la muerte) y por tanto están censurados. Otro ejemplo interesante de datos censurados se encuentra en el intento de estimar las alturas de una población como la norteamericana basándose en la altura de los reclutas de las fuerzas armadas [como se cita en 9]. Los datos están censurados porque durante algunos años las fuerzas armadas tenían restricciones a la altura que varían en nivel y cumplimiento. Por tanto, los investigadores se enfrentan con la tarea de estimar las alturas de toda la población cuando se sabe que ciertos individuos (es decir, aquellos que están por debajo de las restricciones de altura) no están incluidos en las muestras. En ambos casos el conocimiento del investigador del proceso de ausencia de datos permite el uso de métodos especializados, tales como un análisis de la historia del evento para dar cabida a los datos censurados [9].

La justificación para designar a los datos ausentes como prescindibles es que el proceso de ausencia de datos está operando aleatoriamente (es decir, los valores observados son una muestra aleatoria del conjunto total de valores, observados y perdidos) y que esos efectos aleatorios son identificables y explícitamente ajustados a la técnica usada. Sin embargo, en muchos casos, el proceso de ausencia de datos no está explícitamente provocado por la técnica usada. No obstante, el investigador debe evaluar la medida y el impacto en que los datos ausentes determinan si es un proceso aleatorio o, en caso contrario, si se puede remediar con alguna de las soluciones existentes.

Otros tipos de procesos de ausencia de datos

La ausencia de datos puede ocurrir por muchas razones y en muchas situaciones. Un tipo de proceso de datos ausentes que puede ocurrir en cualquier situación se debe a factores de procedimiento, tales como errores en la entrada de datos que crean códigos inválidos, restricciones de representatividad (por ejemplo, los datos de los condados pequeños en el censo de los Estados Unidos), fallos al completar el cuestionario o incluso la morbidez del encuestado. En estas situaciones, el investigador tiene escaso control sobre los procesos de ausencia de datos, aunque pueden aplicarse ciertas soluciones si se encuentra que los datos ausentes son de carácter aleatorio. Otro tipo de procesos de pérdida de datos tiene lugar cuando la respuesta es inaplicable, como las preguntas en relación a los años de matrimonio para adultos que nunca han estado casados. De nuevo, los análisis pueden ser específicamente formulados para acomodar a estos encuestados.

Otros tipos de procesos de datos ausentes se identifican y manejan con menos facilidad. La mayoría de estos tipos están relacionados directamente con el encuestado. Un buen ejemplo es la renuncia del encuestado a responder a ciertas cuestiones. Esto es habitual en cuestiones de naturaleza sensible (tales como aquellas que conciernen a la renta o a asuntos particularmente controvertidos) o cuando el encuestado no tiene opinión o su conocimiento es insuficiente para contestar la pregunta. El investigador debería anticipar estos problemas e intentar minimizarlos en el diseño de la investigación y en los niveles de recogida de datos. Sin embargo, puede ocurrir muy bien que el investigador deba enfrentarse con los datos ausentes resultantes. Pero no todo está perdido. Cuando los datos ausentes ocurren siguiendo una pauta aleatoria, pueden existir soluciones para mitigar sus efectos.

Examen de los tipos de datos ausentes

Para decidir si se puede aplicar una solución para la ausencia de datos, en primer lugar el investigador debe averiguar el grado de aleatoriedad presente en los datos ausentes. Supongamos que se observan dos variables (X e Y). Mientras que X no tiene datos perdidos, Y tiene algunos datos ausentes. Si se encuentra un proceso de datos ausentes entre X e Y, donde existen diferencias significativas para casos de Y con datos válidos y datos ausentes en función de los valores de X, entonces los datos ausentes no son aleatorios. Cualquier análisis tiene que comprobar explícitamente los procesos de datos ausentes entre X y Y o si no se introduce sesgo en los resultados.

Los datos ausentes obedecen a un proceso aleatorio (**MAR**) si los valores ausentes de Y dependen de X, pero no en Y. Con esto queremos decir que los valores observados de Y representan una muestra de los valores reales de Y para cada valor de X, pero los datos observados para Y no representan necesariamente una muestra verdaderamente aleatoria para todos los valores de Y. Aunque el proceso de datos ausentes es aleatorio en la muestra, sus valores no son generalizables para la población. Por ejemplo, supongamos que conocemos el género de los encuestados (la variable X) y éstos son preguntados acerca de la renta familiar (la variable Y). Encontramos que los datos ausentes son aleatorios para ambos sexos pero que ocurren con mayor frecuencia para los hombres que para las mujeres. Mientras que el proceso de ausencia de datos está operando de forma aleatoria, cualquier solución aplicada a los datos ausentes debe tener en cuenta el género de los encuestados porque el género afecta a la distribución definitiva de los valores de las rentas domésticas.

Cuando hay un mayor nivel de aleatoriedad el proceso es completamente aleatorio (**MCAR**). En estos casos, los valores observados de Y son verdaderamente una muestra aleatoria de todos los valores de Y, sin un proceso subyacente que tiende a sesgar los datos observados. En nuestro ejemplo anterior, esto se mostraría por el hecho de que los datos ausentes para las rentas domésticas estén aleatoriamente ausentes en la misma proporción tanto para hombres como para mujeres. Si esta es la forma del proceso de ausencia de datos, cualquier solución se podría aplicar sin tener en cuenta el impacto de cualquier otra variable o proceso de datos ausentes.

Diagnóstico de la aleatoriedad en el proceso de pérdida de observaciones

Como ya se ha apuntado, el investigador debe averiguar si el proceso de ausencia de datos tiene lugar de forma aleatoria. Existen tres métodos para ayudar en este diagnóstico. El primero valora los datos ausentes para una única variable *Y* al formar dos grupos —observaciones con datos ausentes para *Y* y aquellos con valores válidos de *Y*. Se realizan entonces los test para determinar si existen diferencias significativas entre los dos grupos sobre otras variables de interés. Si se encuentran patrones de diferencias significativas, indicaría que existe un proceso de pérdida de datos no aleatorio. Utilicemos otra vez nuestro ejemplo de la renta doméstica y el género. En primer lugar formaríamos dos grupos de encuestados, aquellos con datos ausentes en la pregunta sobre las economías domésticas y aquellos que responden a la pregunta. Podemos entonces comparar los porcentajes de género de cada grupo. Si un género (por ejemplo, hombres) se encontrara en mayor proporción en el grupo de datos ausentes, sospecharíamos que el proceso no ha operado de forma aleatoria. Si la variable que estábamos comparando fuese métrica (por ejemplo, una actitud o percepción) en lugar de categórica (género), entonces el apropiado es el test de la *t*. El investigador debería examinar un número de variables para ver si surge cualquier tipo de patrón consistente. Recordemos que ciertas diferencias pueden ocurrir por azar, pero que cualquier serie de diferencias puede indicar un patrón subyacente.

Una segunda aproximación consistiría en utilizar las correlaciones dicotomizadas para evaluar la correlación de los datos ausentes en cualquier par de valores. Para cada variable, los valores válidos se representan por el valor uno, mientras que los datos ausentes son reemplazados por valores de cero. Estos indicadores de valores perdidos para cada variable estarían entonces correlacionados. Las correlaciones indican el grado de asociación entre los valores perdidos sobre cada par de variables. Bajas correlaciones implican aleatoriedad en el par de variables. Aunque no existen líneas maestras para identificar el nivel de correlación necesario para indicar que los datos ausentes no son aleatorios, los test de significación estadística de las correlaciones proporcionan una estimación conservadora del grado de aleatoriedad. Si la aleatoriedad es indicativa para todos los pares de variables, entonces el investigador puede suponer que los datos ausentes pueden clasificarse como MCAR. Si existen correlaciones significativas entre algunos pares de variables, entonces el investigador puede tener que suponer que los datos son sólo MAR y estas relaciones deben ser tenidas en cuenta en cualquier solución que se quiera aplicar.

Finalmente, se puede realizar un test conjunto de aleatoriedad que determine si los datos ausentes pueden ser clasificados como MCAR. Este test analiza el patrón de datos ausentes sobre todas las variables y las compara con el patrón esperado para un proceso de datos ausentes aleatorio. Si no se encuentran diferencias significativas, los datos ausentes puede ser clasificados como MCAR. Si se encuentran diferencias significativas, sin embargo, el investigador debe usar las aproximaciones descritas más arriba para identificar los procesos específicos de datos ausentes que no son aleatorios.

Aproximaciones al tratamiento de datos ausentes

Las aproximaciones o soluciones que tratan con los datos ausentes pueden clasificarse en cuatro categorías basadas en la aleatoriedad de los procesos de datos ausentes, en función del método empleado para estimarlos [9]. Si se encuentran procesos de datos ausentes MAR o no aleatorios, el investigador debería aplicar sólo el método diseñado específicamente para este proceso [9]. La aplicación de cualquier otro método introduce sesgos en los resultados. Sólo si el investigador determina que el proceso de ausencia de datos puede clasificarse como MCAR pueden utilizarse todas las aproximaciones discutidas en las siguientes secciones.

No obstante, muchas veces los investigadores evalúan la aleatoriedad de los datos ausentes antes de aplicar una de las soluciones de datos ausentes. E incluso si la solución es la apropiada, el

investigador debe tener en cuenta los impactos específicos de los resultados asociados con ella. Demasiadas veces se aplica una solución sin una evaluación de los procesos de ausencia de datos, la conveniencia de la solución seleccionada o las consecuencias que tendrá. En tal caso, el investigador nunca se dará cuenta de los efectos porque están camuflados bajo los resultados generales.

Utilizar sólo aquellas observaciones con datos completos

La aproximación más simple y directa para tratar con los datos ausentes consiste en incluir sólo aquellas observaciones con datos completos, también conocida como la **aproximación de casos completos**. Este método se encuentra en todos los programas estadísticos y es el método por defecto en muchos programas. No obstante, esta aproximación debería usarse sólo si los datos ausentes son MCAR, porque los datos ausentes que no lo son tienen elementos no aleatorios que sesgarían los resultados. Por tanto, incluso aunque sólo se usen observaciones válidas, los resultados no son generalizables para la población. Más aún, en muchas situaciones, el tamaño de la muestra resultante queda reducida a una muestra inapropiada para los propósitos del análisis. La aproximación de casos completos se ajusta mejor a casos en los que la extensión de la ausencia de datos es pequeña, en los que la muestra es suficientemente grande para permitir la supresión de los casos con los datos ausentes y en los que las relaciones entre los datos son tan fuertes que no pueden verse afectadas por cualquier proceso de datos ausentes.

Supresión de caso(s) y/o variable(s)

Otra solución simple para los datos ausentes consiste en suprimir el caso(s) y/o variable(s) que peor se comporta(n) respecto a los datos ausentes. En esta aproximación, el investigador determina la extensión de los datos ausentes sobre cada caso y variable y entonces destruye los casos y variables que exceden el nivel especificado. En muchos casos donde se presenta un patrón de datos no aleatorio, puede constituir la solución más eficiente. El investigador puede encontrar que los datos ausentes están concentrados en un pequeño subconjunto de casos y/o variables, con cuya exclusión se reduce sustancialmente la extensión de los datos ausentes. De nuevo, no existen líneas maestras para el nivel de exclusión necesario, pero cualquier decisión debería basarse tanto en consideraciones empíricas como teóricas. Si se encuentran los valores ausentes para lo que será una variable dependiente en el análisis propuesto, habitualmente se excluye el caso. De esta forma se evita cualquier aumento artificial en el poder explicativo del análisis que pudiera ocurrir cuando el investigador estime en primer lugar los datos ausentes para la variable dependiente por uno de los procesos de imputación descritos a continuación y después use los valores estimados en el análisis de las relaciones de dependencia. Si una variable que no sea la dependiente tiene valores ausentes y es una candidata a la eliminación, el investigador debe asegurarse de que existan variables alternativas, que se espera estén altamente correlacionadas, para representar la intención de la variable original. El investigador debe siempre considerar lo que gana al eliminar una fuente de datos ausentes y lo que pierde al no contar con una determinada variable en el análisis multivariante.

Métodos de imputación

Una tercera categoría para tratar con datos ausentes es a través de uno de los muchos **métodos de imputación**. La imputación es el proceso de estimación de valores ausentes basado en valores válidos de otras variables y/o casos de la muestra. El objetivo es emplear relaciones conocidas que puedan identificarse en los valores válidos de la muestra para ayudar en la estimación de valores ausentes. Sin embargo, el investigador debería considerar cuidadosamente el uso de la imputación en cada instancia, dados sus potenciales impactos sobre el análisis [6]:

> La idea de imputación es tan seductora como peligrosa. Es seductora porque puede llevar al usuario a creer que los datos están completos después de todo, y es peligrosa porque une situaciones donde el problema es suficientemente menor con situaciones donde los estimadores estándar aplicados a los datos reales e imputados tienen sesgos sustanciales.

Los métodos abordados en esta sección suelen utilizarse principalmente con variables métricas por dos razones. En primer lugar, se pueden hacer estimaciones de los datos ausentes para variables métricas con valores como una media de todos los valores válidos. En segundo lugar, las variables no métricas requieren una estimación de un valor específico en vez de una estimación en una escala continua. Existe mucha diferencia entre estimar un valor ausente para una variable métrica, tal como una actitud o una percepción e incluso la renta, que estimar el género del encuestado cuando este dato está ausente. Por tanto, las variables no métricas no se logran típicamente mediante el proceso de imputación, sino que requieren la aproximación de la modelización específica abordada en la siguiente sección o se omiten por estar ausentes.

Se pueden definir los métodos de imputación como pertenecientes a dos tipos: (1) como el uso de toda la información disponible a partir de un subconjunto de casos para generalizar sobre la muestra entera, o (2) como métodos para estimar valores de reemplazo para los datos ausentes que, de esta forma, se analizan mediante técnicas multivariantes estándar. El siguiente debate indicará las diferentes opciones dentro de cada tipo de imputación y sus ventajas y desventajas.

El uso de toda la información disponible como técnica de imputación

El primer tipo de método de imputación no reemplaza los datos ausentes sino que imputa las características de distribución (por ejemplo, la desviación media o estándar) o las relaciones (por ejemplo, correlaciones) de todos los valores válidos disponibles.

Conocido como **enfoque de disponibilidad completa**, este método (la opción PAIRWISE en SPSS y las opciones CORPAIR, COVPAIR o ALLVALUE en BMDP) se usa principalmente para estimar correlaciones y maximizar la información disponible en la muestra. La característica diferencial de esta aproximación es que cada correlación se basa en un conjunto de observaciones potencialmente único y que el número de observaciones empleadas en los cálculos puede variar en cada correlación. El proceso de imputación no consiste en reemplazar los datos ausentes por el resto de los casos, sino en utilizar las correlaciones obtenidas como representantes para la muestra entera. Se puede comparar esta aproximación al enfoque de disponibilidad completa mencionado anteriormente, que usa solamente datos de observaciones que no tienen datos ausentes. Cualquiera de las dos aproximaciones puede introducir sesgos si el proceso de datos ausentes no es MCAR.

Aunque el método de disponibilidad completa maximiza los datos utilizados y salva el problema de los datos ausentes de una única variable eliminando un caso del análisis entero, pueden también surgir muchos problemas de esta aproximación. En primer lugar, las correlaciones pueden calcularse «fuera de rango» y de forma inconsistente con otras correlaciones de la matriz de correlación. Cualquier correlación entre X e Y queda restringida por su correlación con una tercera variable Z, tal y como indica la siguiente fórmula:

$$\text{Rango de } r_{xy} = r_{xz}\, r_{yz} \pm \sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}$$

La correlación entre X e Y puede variar sólo de +1 a –1 si tanto X como Y tienen una correlación cero con todas las otras variables de la matriz de correlación. Pero rara vez se dan correlaciones con otras variables distintas de cero. En la medida en que las correlaciones con otras variables aumenten, el rango de posibles correlaciones entre X e Y disminuye, aumentando de esta forma la posibilidad de que la correlación en un único conjunto de casos sea inconsistente con las correlaciones derivadas de otros conjuntos de casos. Por ejemplo, si X y Y tienen co-

rrelaciones de 0,6 y 0,4 respectivamente con Z, entonces el rango de correlaciones posibles entre X e Y es 0,24 $\pm$ 0,73, o de –0,49 a 0,97. Cualquier valor fuera de este rango es matemáticamente inconsistente, aunque podría ocurrir si se obtiene la correlación con un número y conjunto de casos diferentes para las dos correlaciones en el enfoque de disponibilidad completa.

Un problema asociado es que los autovalores de la matriz de correlación pueden llegar a ser negativos, alterando así las propiedades de varianza de la matriz de correlación. Aunque la matriz de correlación puede ajustarse para eliminar este problema (por ejemplo, la opción ALLVALUE en BMDP), muchos programas no incluyen este programa de ajuste. En casos extremos, la matriz estimada de varianzas/covarianzas no es positiva definida. Todos estos problemas deben ser considerados al seleccionar esta aproximación, frente a excluir casos con datos ausentes.

Sustitución de datos ausentes

La segunda forma de imputación consiste en el método efectivo de sustitución de los datos ausentes por valores estimados sobre la base de otra información existente en la muestra. Esta medida puede llevarse a cabo de muchas maneras, que van desde una sustitución directa de valores, a procesos de estimación basados en relaciones entre variables. La exposición siguiente se centrará en los métodos más ampliamente utilizados, aunque existen otras muchas formas de imputación [9].

Sustitución de caso En este método, las observaciones con datos ausentes se sustituyen con otras observaciones no muestrales. Un ejemplo común es reemplazar un hogar que está en la muestra pero con el que no se puede contactar o que tiene gran cantidad de datos ausentes con otro hogar que no está en la muestra, preferiblemente muy similar al de la observación original.

Este método es el que más se utiliza para sustituir las observaciones con datos ausentes completos, aunque también puede emplearse para reemplazar observaciones con menores cantidades de datos ausentes.

Sustitución por la media Uno de los métodos más empleados consiste en sustituir los valores ausentes por una variable cuyo valor medio se calcula sobre todas las respuestas válidas. De esta forma, las respuestas de la muestra válida se usan para calcular el valor de sustitución. La lógica de esta aproximación es que la media es el mejor valor de sustitución. Esta aproximación, aunque es extensamente utilizada, tiene tres desventajas. En primer lugar, invalida las estimaciones de la varianza derivadas de las fórmulas estándar de la varianza para conocer la verdadera varianza de los datos. En segundo lugar, la distribución real de los valores se encuentra distorsionada por la sustitución de los datos ausentes por la media. En tercer lugar, este método modifica la correlación observada porque todos los datos ausentes tendrán un valor único constante. Sin embargo, tiene la ventaja de que se puede llevar a cabo fácilmente y de proporcionar una información completa para todos los casos.

Sustitución por valor constante En este método, el investigador sustituye los datos ausentes por un valor constante derivado de fuentes externas o investigación previa. Su naturaleza es similar al método de sustitución de la media, que difiere sólo en la fuente del valor de sustitución. La imputación de valor constante tiene las mismas desventajas que el método de sustitución de la media, y el investigador debe asegurarse que el valor de sustitución de una fuente externa es más válido que el valor generado internamente por la media. Este método puede proporcionar la opción de reemplazar los datos ausentes con un valor que podría ser considerado más válido que la media de la muestra.

Imputación por regresión En este método (tal y como se indica en el Capítulo 4) se usa el análisis de regresión para predecir los valores ausentes de una variable basándose en su relación con otras variables del conjunto de datos. Al mismo tiempo que tiene el atractivo de usar las relaciones ya existentes en la muestra como base de predicción, también tenemos varias desventajas asociadas con este método. En primer lugar, refuerza las relaciones ya existentes en los datos. Conforme aumente el uso de este método, los datos resultantes son más característicos de la muestra y menos generalizable. En segundo lugar, a menos que se añadan valores estocásticos a los valores estimados, se subestima la varianza de la distribución. En tercer lugar, este método supone que la variable con datos ausentes tiene correlaciones sustanciales con otras variables. Si estas correlaciones no son suficientes para producir una estimación significativa, entonces son preferibles otros métodos, como la sustitución por la media. Finalmente, el procedimiento de regresión no está restringido en las estimaciones que hace. Por tanto, los valores predichos puede que no correspondan a los rangos válidos de las variables (por ejemplo, predecir un valor de 11 para una escala de 10 puntos), requiriendo por tanto alguna forma de ajuste adicional. Incluso con todos estos problemas potenciales, el método de regresión de imputación es prometedor en aquellos casos donde se presenten niveles moderados de dispersión de los datos ausentes y donde las relaciones entre las variables están lo suficientemente establecidas como para que el investigador confíe en que el uso del método no tendrá impacto sobre la generalidad de los resultados.

Imputación múltiple El último método de imputación es en realidad una combinación de varios métodos. En esta aproximación, se usan dos o más métodos para derivar una estimación compuesta, usualmente la media de las diversas estimaciones para el dato ausente. La lógica de esta aproximación es que el uso de la aproximación múltiple minimiza los problemas específicos con cualquier método simple siendo su composición la mejor estimación. La elección de esta aproximación se basa fundamentalmente en la concesión mutua entre la percepción del investigador de los potenciales beneficios ponderada y el esfuerzo sustancialmente superior que requiere realizar y combinar las múltiples estimaciones.

Procedimientos basados en el modelo

Este conjunto final de procedimientos incorpora explícitamente los datos ausentes en el análisis, bien sea a través de un proceso específicamente diseñado para la estimación de datos ausentes, o bien como una porción integral del análisis multivariante estándar. La primera aproximación utiliza estimaciones de máxima verosimilitud que intentan modelizar los procesos que subyacen en la ausencia de datos y realizar la estimación más precisa y razonable [9]. Un ejemplo es la aproximación EM en SPSS. Representa un método frecuente de dos etapas (las etapas E y M) en los que la etapa E realiza las mejores posibles estimaciones de los datos ausentes y a continuación la etapa M realiza estimaciones de los parámetros (medias, desviaciones típicas o correlaciones) con la suposición de que se reemplazaron todos los datos ausentes. El proceso continúa con estas dos etapas hasta que el cambio de los valores estimados es despreciable y se reemplazan todos los datos ausentes.

La segunda aproximación incluye los datos ausentes directamente en el análisis, definiendo observaciones con datos ausentes como un subconjunto selecto de la muestra. Esta aproximación es más apropiada para tratar con los datos ausentes de las variables independientes cunado hay una relación de dependencia. Su premisa queda caracterizada perfectamente en este pasaje de Cohen y Cohen [4, pág. 299]:

> Vemos por tanto los datos ausentes como un hecho práctico que debe ser investigado, en lugar de como un desastre a ser mitigado. Además, implícita en esta filosofía está la idea de que como otros aspectos de la muestra de datos, los datos ausentes son una propiedad de la población a la que se busca generalizar.

Cuando tenemos datos ausentes en una variable no métrica, el investigador puede definir fácilmente aquellas observaciones como un grupo separado y entonces incluirlas en cualquier análisis, como ANOVA o MANOVA o incluso el análisis discriminante. Cuando los datos ausentes se encuentran en una variable independiente métrica de una relación de dependencia, se ha desarrollado un procedimiento para incorporar las observaciones en el análisis mientras se mantienen las relaciones entre las variables válidas [4]. Este procedimiento se puede ilustrar mejor en el contexto de un análisis de regresión, aunque también puede utilizarse en otras relaciones de dependencia. El primer paso es codificar todas las observaciones con datos ausentes como variables ficticias (donde los casos con datos ausentes reciben un valor de uno y el resto de los casos un valor de cero). Los valores ausentes se imputan así por el método de la sustitución por la media. Finalmente, la relación se estima por medias normales. Las variables ficticias representan la diferencia en la variable dependiente entre aquellas observaciones con datos ausentes y aquellas observaciones con datos válidos. El test del coeficiente de la variable ficticia evalúa la significación estadística de esta diferencia. El coeficiente de la variable original representa la relación entre todos los casos con datos no ausentes. Este método permite al analista retener todas las observaciones en el análisis con el fin de mantener el tamaño de la muestra, mientras que también proporciona un test directo de las diferencias entre los dos grupos junto con las relaciones estimadas entre variables dependientes e independientes.

Un ejemplo de diagnóstico de datos ausentes

Para ilustrar el proceso de diagnóstico de las pautas de datos ausentes y la aplicación de las posibles soluciones, vamos a introducir un nuevo conjunto de datos (veáse Apéndice A para una lista completa de las observaciones). Este conjunto de datos fue recogido durante unas pruebas previas del cuestionario utilizado para recoger los datos descritos en el Capítulo 1. Esta prueba previa se realizó a 70 individuos y recogía respuestas sobre las 14 variables. En el curso de esta prueba, sin embargo, se encontraron datos ausentes. Las siguientes secciones detallan el diagnóstico de la extensión de los datos ausentes en el conjunto de datos y los análisis proporcionados para seleccionar y aplicar las diversas soluciones para la ausencia de datos disponibles en la mayor parte de los programas estadísticos. Muchos programas de software están incorporando programas de análisis de datos ausentes, entre ellos BMDP y SPSS. Los análisis indicados a continuación pueden ser reproducidos mediante la manipulación de datos y el análisis convencional. En el Apéndice A se proporcionan varios ejemplos.

Examen de las pautas de datos ausentes

La Tabla 2.2 contiene las estadísticas descriptivas de las observaciones con datos válidos, incluyendo el porcentaje de casos con datos ausentes sobre cada variable. Se han eliminado seis casos del análisis debido a que se trataba esencialmente de datos ausentes en todas las variables de interés. La cantidad de datos ausentes para las 64 observaciones restantes va desde el máximo de un 30 por ciento de los casos para X_1 a un mínimo de un único caso (1,6 por ciento) para X_6. Para las variables con un mayor nivel de datos ausentes (X_1, X_2 y X_3), los niveles no son tan excesivos para dictar una exclusión de la variable. Dado el papel integral que se espera que jueguen estas variables en los diversos análisis multivariantes, se deben realizar todos los esfuerzos posibles para retenerlos en el análisis.

Estadística de resumen de los datos previos a la prueba

Un factor que podría aliviar algunos de los altos niveles de datos ausentes para ciertas variables es la eliminación de los casos del análisis. Para determinar si los datos ausentes se con-

TABLA 2.2. Estadística de resumen de los datos previos a la prueba

	Número de casos con datos válidos	*Media*	*Desviación estándar*	*Datos ausentes*	
				Número	*Porcentaje*
X_1 Velocidad de entrega	45	4,0133	0,9664	19	29,7
X_2 Nivel de precios	54	1,8963	0,8589	10	15,6
X_3 Flexibilidad de precios	50	8,1300	1,3194	14	21,9
X_4 Imagen del productor	60	5,1467	1,1877	4	6,3
X_5 Servicio conjunto	59	2,8390	0,7541	5	7,8
X_6 Imagen de fuerza de ventas	63	2,6016	0,7192	1	1,6
X_7 Calidad de producto	60	6,7900	1,6751	4	6,3
X_9 Nivel de fidelidad	60	45,9667	9,4204	4	6,3
X_{10} Nivel de satisfacción	60	4,7983	0,8194	4	6,3

Nota: Seis de los 70 casos originales tenían más del 90 por ciento de datos ausentes y fueron excluidos del análisis. Todos los análisis se basan en los 64 casos restantes. Veintiseis casos no tenían datos ausentes.

centran en un conjunto seleccionado de casos, la Tabla 2.3 proporciona una representación gráfica de las pautas de datos ausentes. Excepto los seis casos ya eliminados debido a los niveles extremadamente altos de datos ausentes, vemos que ningún otro caso tiene un número desproporcionado de datos ausentes. De hecho, sólo dos casos tienen más de dos datos ausentes. Por tanto, de los 38 casos con datos ausentes, solamente cuatro casos tienen más de dos valores ausentes.

TABLA 2.3. Representación gráfica de las pautas de datos ausentes

		Clase de caso								
		Datos ausentes								
Caso	*Número de variables ausentes*	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_9	X_{10}
202	2	S		S						
203	2		S					S		
204	3	S		S						S
205	1			S						
207	3	S		S						S
213	2		S	S						
216	2	S				S				
218	2	S				S				
219	2							S	S	
220	1		S							
221	3	S		S				S		
222	2			S		S				
224	3	S	S						S	
225	2			S	S					
227	2		S						S	
228	2	S			S					
229	1					S				
231	1							S		
232	2	S	S							
235	2						S			S
237	1		S							
238	1	S								
240	1	S								

TABLA 2.3. (*continuación*) Representación gráfica de las pautas de datos ausentes

		Clase de caso								
		Datos ausentes								
Caso	*Número de valores ausentes*	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_9	X_{10}
241	2			S		S				
244	1								S	
246	1				S					
248	2	S	S							
249	1		S							
250	2	S		S						
253	1	S								
255	2	S		S						
256	1	S								
257	2		S	S						
259	1	S								
260	1	S								
267	2			S	S					S
268	1									
269	2	S		S						

Leyenda: S = un valor ausente.

La Tabla 2.4 representa las pautas de los datos ausentes. La pauta más frecuente es la que se encuentra en los datos ausentes de X_1 (seis casos), y la siguiente pauta más común está en los da-

TABLA 2.4. Pautas de datos ausentes tabuladas

	Pautas de datos ausentes[a]								
Número de casos	X_6	X_{10}	X_4	X_7	X_9	X_5	X_2	X_3	X_1
26								X	
1								X	X
4									X
6									X
1			X						
1			X						
2			X					X	
2						X		X	
1						X			
2						X			X
2							X		X
3							X		
2							X	X	
1				X			X		
1				X					
1				X	X				
1					X				
1					X		X		
1					X		X		X
1		X							
1	X	X						X	X
2		X						X	X
1				X					

[a] Las variables están ordenadas sobre las pautas ausentes.

tos ausentes de X_1 y X_3 (cuatro casos). Todos los casos restantes muestran pautas que son esencialmente únicas o compartidas solamente en un número de casos muy reducido. Tal y como demuestra este análisis, no ocurren pautas con una frecuencia que sugiere un proceso de datos ausentes subyacente. Por tanto, no se puede eliminar ningún caso o serie de casos.

La diagnosis de la aleatoriedad de los datos ausentes

El siguiente paso es examinar las pautas de datos ausentes y determinar si los datos ausentes se distribuyen aleatoriamente a través de los casos y de las variables. El primer test para evaluar la aleatoriedad es comparar las observaciones con y sin datos ausentes para cada variable en función de las otras variables. Por ejemplo, comprobar si las observaciones con datos ausentes en X_1 se sitúan en un grupo y las observaciones con respuestas válidas en X_1 se sitúan en otro grupo. A continuación, estos dos grupos se comparan para identificar cualquier diferencia entre las variables restantes (de X_2 a X_{10}). Una vez que se han hecho las comparaciones sobre las otras variables, se forman nuevos grupos sobre los datos ausentes para la siguiente variable (X_2) y se comparan otra vez sobre el resto de las variables. Este proceso continúa hasta que cada variable (desde X_1 a X_{10}) ha sido examinada para cada diferencia. El objetivo es identificar cualquier proceso sistemático de datos ausentes que quede reflejado en pautas de diferencias significativas.

La Tabla 2.5 contiene los resultados de este análisis para las 64 observaciones restantes de la muestra antes de la prueba. La primera pauta notable de valores *t* significativos se observa en X_9, donde se han encontrado diferencias significativas entre los dos grupos en seis de las nueve comparaciones. Sin embargo, el impacto de estas diferencias es marginal dado que el número de casos con datos ausentes en X_9 oscila entre tres y cinco. X_7 muestra una pauta de diferencias similar a X_9 con sólo cuatro diferencias significativas y pequeños grupos de casos con datos ausentes. Este análisis indica que aunque se pueden encontrar diferencias significativas debido a los datos ausentes de dos variables (X_7 y X_9), sin embargo, el reducido número de casos implicados hace que el interés sea marginal. Si los test de aleatoriedad ulteriores indican una pauta no aleatoria de los datos ausentes, estos resultados proporcionarían entonces un punto de partida para posibles soluciones. Un segundo test de aleatoriedad implica el uso de correlaciones entre variables dicotómicas. Las

TABLA 2.5. Evaluación de la aleatoriedad de los datos ausentes a través de comparaciones de grupos de observaciones con datos ausentes frente a datos válidos

Grupos formados por datos ausentes sobre:		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_9	X_{10}
X_1	*t*		–0,3	1,3	2,2	2,6	1,9	–1,1	2,6	2,1
	Significación		0,763	0,223	0,033	0,017	0,065	0,273	0,017	0,049
	Número presente	45	38	38	42	42	44	42	42	43
	Número ausente	0	16	12	18	17	19	18	18	17
	Media (presente)	4,01	1,87	8,27	5,34	3,02	2,71	6,61	48,17	4,95
	Media (ausente)		1,95	7,68	4,69	2,39	2,36	7,20	40,83	4,42
X_2	*t*	–0,5		0,7	–2,2	–4,2	–2,4	–1,2	–1,1	–1,2
	Significación	0,646		0,528	0,044	0,001	0,034	0,260	0,318	0,233
	Número presente	38	54	42	50	49	53	51	52	50
	Número ausente	7	0	8	10	10	10	9	8	10
	Media (presente)	3,97	1,90	8,18	4,99	2,70	2,51	6,68	45,46	4,75
	Media (ausente)	4,23		7,86	5,94	3,50	3,11	7,40	49,25	5,02
X_3	*t*	0,4	1,4		1,1	2,0	0,2	0,0	1,9	0,9
	Significación	0,693	0,180		0,286	0,066	0,818	0,965	0,073	0,399
	Número presente	38	42	50	48	47	49	47	46	48
	Número ausente	7	12	0	12	12	14	13	14	12
	Media (presente)	4,03	1,98	8,13	5,24	2,95	2,61	6,80	47,02	4,84
	Media (ausente)	3,90	1,60		4,79	2,42	2,56	6,77	42,50	4,62

TABLA 2.5. (*continuación*) Evaluación de la aleatoriedad de los datos ausentes a través de comparaciones de grupos de observaciones con datos ausentes frente a datos válidos

Grupos formados por datos ausentes sobre:		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_9	X_{10}
X_4	*t*	–0,2	2,6	–0,3		0,2	1,4	1,5	0,2	–2,4
	Significación	0,882	0,046	0,785		0,888	0,249	0,197	0,830	0,064
	Número presente	42	50	48	60	55	59	56	56	56
	Número ausente	3	4	2	0	4	4	4	4	4
	Media (presente)	4,01	1,94	8,12	5,15	2,84	2,63	6,83	46,02	4,76
	Media (ausente)	4,07	1,33	8,35		2,80	2,25	6,20	45,25	5,38
X_5	*t*	–0,1	–0,3	0,8	0,4		–0,9	–0,4	0,5	0,6
	Significación	0,900	0,749	0,502	0,734		0,423	0,696	0,669	0,605
	Número presente	42	49	47	55	59	58	55	55	55
	Número ausente	3	5	3	5	0	5	5	5	5
	Media (presente)	4,01	1,89	8,20	5,16	2,84	2,58	6,76	46,18	4,82
	Media (ausente)	4,10	1,98	7,10	5,04		2,86	7,14	43,60	4,56
X_7	*t*	3,0	0,9	0,2	–2,1	0,9	–1,5		0,5	0,4
	Significación	0,036	0,440	0,864	0,118	0,441	0,193		0,658	0.704
	Número presente	42	51	47	56	55	59	60	57	56
	Número ausente	3	3	3	4	4	4	0	3	4
	Media (presente)	4,07	1,92	8,14	5,07	2,86	2,58	6,79	46,14	4,81
	Media (ausente)	3,27	1,50	8,00	6,18	2,55	2,90		42,67	4,70
X_9	t	6,1	–1,4	2,2	–1,1	–0,9	–1,8	1,7		1,6
	Significación	0,000	0,384	0,101	0,326	0,401	0,149	0,128		0,155
	Número presente	42	52	46	56	55	59	57	60	56
	Número ausente	3	2	4	4	4	4	3	0	4
	Media (presente)	4,08	1,85	8,26	5,11	2,82	2,57	6,82	45,97	4,82
	Media (ausente)	3,10	3,00	6,63	5,62	3,08	3,03	6,30		4,47
X_{10}	*t*	1,7	0,8	–2,1	2,5	2,7	1,3	0,9	2,4	
	Significación	0,249	0,463	0,235	0,076	0,056	0,302	0,409	0,066	
	Número presente	43	50	48	56	55	60	56	56	60
	Número ausente	2	4	2	4	4	3	4	4	0
	Media (presente)	4,03	1,92	8,09	5,23	2,89	2,62	6,83	46,43	4,80
	Media (ausente)	3,55	1,60	9,20	3,95	2,08	2,17	6,30	39,50	

Cada celda contiene seis valores: 1) valor *t* para comparar la variable de columna entre un grupo a (observaciones con datos válidos sobre la variable lineal) y grupo b (observaciones con datos ausentes sobre la variable lineal); 2) significación de valor *t* para comparaciones de grupo; 3) y 4) número de observaciones para el grupo a (datos válidos) y grupo b (datos ausentes); 5) y 6) media de variable para el grupo a (datos válidos) y grupo b (datos ausentes).

Interpretación de la tabla:

La celda de la derecha superior indica que el valor *t* para la comparación de X_{10} entre el grupo a (datos válidos) y grupo b (datos ausentes) sobre X_1 es 2,1, lo que tiene un nivel de significación de 0,049. Los tamaños muestrales del grupo a y el grupo b son 43 y 17, respectivamente. Finalmente, la media de X_{10} para el grupo a (datos válidos) es 4,95, mientras que la media para el grupo b (datos ausentes) es 4,42.

variables dicotómicas se forman al reemplazar los valores válidos por uno y los datos ausentes por cero. Las correlaciones resultantes entre las variables dicotómicas indican la medida en que los datos ausentes están relacionados entre pares de variables. Las correlaciones bajas indican una baja asociación entre los procesos de ausencia de datos para esas dos variables. La Tabla 2.6 contiene las correlaciones entre las nueve variables dicotomizadas. La revisión de los valores indica que sólo una correlación tiene un rango moderado (X_6 y X_{10} presentan una correlación de 0,488). Esto sugiere que los procesos de datos ausentes que influyen en X_{10} corresponden a los procesos de datos ausentes que afectan a X_6. Sin embargo, dada la ausencia de cualquier otra correlación con valores incluso moderados, el investigador puede asegurarse que ningún proceso aislado de datos ausentes está afectando significativamente a un número sustancial de variables.

TABLA 2.6. Evaluación de la aleatoriedad de los datos ausentes a través de correlaciones de una variable dicotomizada y el test multivariante para ausencia complemente aleatoria (MCAR)

	X_1 *Velocidad de entrega*	X_2 *Nivel de precios*	X_3 *Flexibilidad de precios*	X_4 *Imagen del productor*	X_5 *Servicio conjunto*	X_6 *Imagen de fuerza de ventas*	X_7 *Calidad de producto*	X_9 *Nivel de fidelidad*	X_{10} *Nivel de satisfacción*
X_1	1,000								
	45								
X_2	0,003	1,000							
	38	54							
X_3	0,235	–0,020	1,000						
	38	42	50						
X_4	–0,026	–0,111	0,176	1,000					
	42	50	48	60					
X_5	0,066	–0,125	0,128	–0,075	1,000				
	42	49	47	55	59				
X_6	–0,082	–0,054	–0,067	–0,033	–0,037	1,000			
	44	53	49	59	58	63			
X_7	–0,026	0,067	0,020	–0,067	–0,075	–0,033	1,000		
	42	51	47	56	55	59	60		
X_9	–0,026	0,244	–0,137	–0,067	–0,075	–0,033	0,200	1,000	
	42	52	46	56	55	59	57	60	
X_{10}	0,115	–0,111	0,176	–0,067	–0,075	0,488*	–0,067	–0,067	1,000
	43	50	48	56	55	60	56	56	60

Test MCAR de Little: Chi-cuadrado: 174,464
Grados de libertad: 159
Probabilidad: 0,190

Interpretación:
El primer valor en la tabla representa la correlación entre las variables dicotomizadas, donde los casos con un valor válido reciben un 1 y los datos ausentes reciben un 0. El segundo valor por debajo de la correlación representa el número de casos que tienen datos válidos sobre las dos variables en esta pareja de correlación específica.

* Significante al nivel de 0,05.

La prueba final es un test conjunto de los datos ausentes para comprobar que la ausencia de datos es completamente aleatoria (MCAR). El test hace una comparación de los datos ausentes con lo que se esperaría si los datos ausentes se distribuyesen de una forma completamente aleatoria. En este caso, como se muestra en la Tabla 2.6, el nivel de significación de los test MCAR es de 0,190, indicando que los procesos de ausencia de datos pueden ser considerados como MCAR. Como resultado, el investigador puede emplear cualquiera de las soluciones para la ausencia de datos, porque no existen sesgos potenciales en las pautas que siguen los datos ausentes.

Soluciones para la ausencia de datos

Como se ha discutido previamente, existen numerosas soluciones para tratar con la ausencia de los datos. Pero en este caso, muchas de ellas tienen claras desventajas. Si se considera la aproximación de caso completo, el tamaño de la muestra se reduce a 26 observaciones, apenas suficiente incluso para el análisis univariante más simple, y mucho menos para aplicaciones multivariantes. Nuestro anterior examen de las pautas de los datos ausentes demostró que no existía un conjunto pequeño de casos que pudieran ser destruidos y así reducir sustancialmente la cuantía de los datos ausentes. Además, la única alternativa viable es la eliminación de X_1, que tiene datos ausentes en casi un 30 por ciento de los casos. Pero incluso si X_1 fuera eliminada, todos los casos con datos ausentes todavía tendrían al menos otra variable con datos ausentes; por tanto, esta alternativa es la aproximación menos atractiva como medio de crear más observaciones con datos completos en todas las variables.

La opción que nos resta es emplear alguna forma de imputación para estimar los valores que reemplazarán a los valores ausentes. Al seleccionar una técnica de imputación, el primer paso es seleccionar entre usar sólo observaciones con datos completos o usar todas la información disponible. La ventaja de la aproximación de información completa es que mantiene la consistencia de la matriz de correlación; pero puede también reducir el número de observaciones utilizado a un conjunto tan pequeño de la muestra (28 casos), que las correlaciones resultantes utilizadas para la imputación difieran significativamente de aquellas observaciones obtenidas de las aproximaciones de información y disponibilidad completa.

La aproximación de disponibilidad completa maximiza el número de casos usados en el cálculo de correlaciones. Una tercera opción es el uso de una sustitución de media para todos los datos ausentes y después calcular las correlaciones.

La Tabla 2.7 contiene las correlaciones obtenidas de las aproximaciones de disponibilidad e información completa y sustitución de la media. En muchos casos las correlaciones son muy similares, pero existen numerosas pautas de diferencias sustanciales. En primer lugar, hay una consistencia entre las correlaciones obtenidas con las aproximaciones de disponibilidad completa y la sustitución por la media. Las diferencias surgen entre las correlaciones obtenidas con la apro-

TABLA 2.7. Comparación de las correlaciones obtenidas con las aproximaciones de disponibilidad completa (en términos de pareja), de caso completo (en términos de lista) y sustitución de la media

	X_1 *Velocidad de entrega*	X_2 *Nivel de precios*	X_3 *Flexibilidad de precios*	X_4 *Imagen del productor*	X_5 *Servicio conjunto*	X_6 *Imagen de fuerza de ventas*	X_7 *Calidad de producto*	X_9 *Nivel de fidelidad*	X_{10} *Nivel de satisfacción*
X_1	1,000								
	1,000								
	1,000								
X_2	–0,479	1,000							
	–0,502	1,000							
	–0,349	1,000							
X_3	0,416	–0,357	1,000						
	0,429	–0,294	1,000						
	0,329	–0,289	1,000						
X_4	–0,099	0,299	–0,065	1,000					
	–0,245	0,320	–0,061	1,000					
	–0,086	0,245	–0,057	1,000					
X_5	0,366	0,440	0,047	0,432	1,000				
	0,566	0,421	0,157	0,046	1,000				
	0,232	0,382	0,042	0,422	1,000				
X_6	0,031	0,260	–0,035	0,810	0,344	1,000			
	–0,094	0,356	–0,066	0,804	0,213	1,000			
	0,027	0,219	–0,032	0,769	0,323	1,000			
X_7	–0,138	0,348	–0,358	0,398	0,066	0,402	1,000		
	–0,416	0,354	–0,230	0,382	–0,150	0,529	1,000		
	–0,106	0,310	–0,297	0,374	0,061	0,395	1,000		
X_9	0,376	0,149	0,601	0,223	0,712	0,268	–0,202	1,000	
	0,599	0,048	0,648	0,191	0,683	0,301	–0,099	1,000	
	0,265	0,134	0,503	0,216	0,656	0,260	–0,195	1,000	
X_{10}	0,514	–0,184	0,702	0,378	0,533	0,233	–0,256	0,669	1,000
	0,549	–0,278	0,725	0,170	0,304	0,064	–0,405	0,566	1,000
	0,381	–0,173	0,626	0,344	0,477	0,229	–0,250	0,647	1,000

Interpretación: El valor superior es la correlación obtenida con una aproximación en términos de pareja o disponibilidad completa, el segundo valor es la correlación obtenida con una aproximación en términos de lista o información completa, y el tercer valor es la correlación obtenida con la sustitución de media. Los tamaños muestrales para la aproximación de la disponibilidad de información completa variaban; los tamaños muestrales reales están listados en la Tabla 2.5. Se utilizó un tamaño muestral de 26 para las correlaciones de información completa; no había datos ausentes después de la sustitución de la media, por lo que el tamaño muestral para esta aproximación fue 64.

TABLA 2.8. Resultados de los métodos de regresión e imputación EM

Media estimada

Métodos de imputación	X_1 *Velocidad de entrega*	X_2 *Nivel de precios*	X_3 *Flexibilidad de precios*	X_4 *Imagen del productor*	X_5 *Servicio conjunto*	X_6 *Imagen de fuerza de ventas*	X_7 *Calidad de producto*	X_9 *Nivel de fidelidad*	X_{10} *Nivel de satisfacción*
EM	3,71	2,03	8,11	5,15	2,82	2,60	6,84	45,85	4,77
Regresión	3,84	1,96	8,10	5,15	2,81	2,59	6,88	45,77	4,77

Desviaciones estándares estimadas

Métodos de imputación	X_1 *Velocidad de entrega*	X_2 *Nivel de precios*	X_3 *Flexibilidad de precios*	X_4 *Imagen del productor*	X_5 *Servicio conjunto*	X_6 *Imagen de fuerza de ventas*	X_7 *Calidad de producto*	X_9 *Nivel de fidelidad*	X_{10} *Nivel de satisfacción*
EM	1,15	1,00	1,27	1,16	0,75	0,71	1,68	9,29	0,82
Regresión	0,99	0,83	1,26	1,15	0,75	0,72	1,69	9,18	0,82

ximación de información completa. En segundo lugar, pueden observarse diferencias notables en las correlaciones entre X_1 y X_{10} con X_4, X_5, X_6 y X_7. Estas diferencias pueden indicar el impacto del proceso de datos ausentes, que no se ha detectado en las medidas de diagnosis previas. Aunque el investigador no tenga una prueba de mayor validez para cada aproximación, estos resultados demuestran las marcadas diferencias que se obtienen a veces entre dos aproximaciones. Cualquiera que sea la aproximación que se elija, el investigador debería examinar las correlaciones obtenidas por métodos alternativos para entender el rango de valores posibles.

Se puede elegir también el método de imputación para estimar los valores que van a reemplazar a los datos ausentes. La Tabla 2.8 contiene algunos resultados obtenidos al emplear la aproximación de regresión de sustitución de la media y EM para la imputación de valores ausentes. Estos resultados incluyen las desviaciones de media y estándar que se obtienen después de reemplazar los valores ausentes por los datos imputados. Tal y como hemos observado en comparaciones de correlaciones anteriores, se pueden detectar algunas diferencias, pero no existe una pauta consistente. Para las variables X_1 y X_2, existen diferencias notables en los valores estimados. Para las variables restantes los resultados son en general muy similares e incluso idénticos. Por tanto, el investigador no tiene que tener una indicación definitiva de cuál de las aproximaciones es apropiada. En su lugar, tiene que unir las pautas de datos ausentes con los puntos fuertes y débiles para cada aproximación y después seleccionar el método más apropiado. En el caso de estimaciones diferentes, el método más apropiado podría ser la aproximación más conservadora de combinar las estimaciones en una única estimación (la aproximación de imputación múltiple). Sea cual sea la aproximación que se usa, el conjunto de datos con valores de reemplazo se debería conservar para análisis posteriores.

Un repaso del análisis de valores ausentes

Nuestra evaluación de las cuestiones que rodean los datos ausentes en los datos anteriores a la prueba puede ser resumida mediante cuatro conclusiones:

1. *El proceso de datos ausentes es MCAR.* Todas las técnicas de diagnosis respaldan la conclusión de que no existe un proceso de datos ausentes sistemático, que convierte los datos

ausentes en MCAR (proceso de datos ausentes completamente aleatorio). Este indicio proporciona al investigador dos ventajas. En primer lugar, no debería haber un impacto «oculto» sobre los resultados que haya que considerar a la hora de interpretar los resultados. En segundo lugar, se puede aplicar cualquiera de los métodos de imputación como solución para los datos ausentes. Su selección no debe basarse en su capacidad para tratar con procesos no aleatorios, sino en la capacidad de aplicar el proceso y su impacto sobre los resultados.

2. *La imputación es la forma de actuar más lógica.* Dado el beneficio mínimo de la supresión de casos y variables, el investigador debe excluir la solución más simple (suprimir casos o variables). Además, el método del caso completo tendría como resultado un tamaño de muestra inadecuado. Por tanto, es necesario llevar a cabo una imputación con el fin de mantener un tamaño de muestra adecuado para cualquier análisis multivariante.
3. *Las correlaciones imputadas son diferentes en todas las técnicas.* Cuando se estiman las correlaciones entre variables en presencia de datos ausentes, el investigador puede elegir entre tres técnicas comunes: el método de información completa, el método de información con disponibilidad completa y el método de sustitución de la media. No obstante, en esta situación se enfrenta con diferencias en los resultados entre estos tres métodos. Las aproximaciones de información de disponibilidad completa y de sustitución de la media conducen, por regla general, a resultados consistentes, aunque los valores de sustitución de la media son generalmente algo más bajos. Hay diferencias notables entre estas dos aproximaciones y la aproximación de información completa. Mientras que la aproximación de información completa parece ser el método más «seguro» y conservador, en este caso no es recomendable dado el tamaño reducido de la muestra utilizada (solamente 26 observaciones) y sus diferencias destacadas con los otros dos métodos. Si es necesario, el investigador debe elegir entre las dos aproximaciones alternativas.
4. *Los métodos múltiples para reemplazar los datos ausentes están disponibles y son apropiados.* Tal y como se ha mencionado anteriormente, la sustitución de la media es una de las maneras aceptables de generar valores de sustitución para los datos ausentes. Además el investigador tiene a su disposición los métodos de regresión y imputación EM, cada uno de los cuales dan estimaciones consistentes para la mayoría de las variables. La presencia de tres métodos aceptables también permite al investigador combinar las tres estimaciones en un compuesto, con la esperanza de mitigar cualquier efecto que se deba estrictamente a uno de estos métodos.

En conclusión, las herramientas y los procesos de diagnosis que se han presentado en esta sección han proporcionado una base adecuada para comprender y acomodar los datos asusentes. Tal y como demuestra este ejemplo, el investigador no debería temer que la presencia de datos ausentes imposibilite la aplicación de un análisis multivariante o limite su capacidad para generalizar los resultados. En su lugar, se puede identificar el posible impacto «oculto» de los datos ausentes y realizar las acciones necesarias para minimizar su efecto sobre los análisis.

Resumen

Los procedimientos disponibles para tratar con datos ausentes varían en forma, complejidad e intención. El investigador debe estar siempre preparado para evaluar y tratar con los datos ausentes que se encontrará con frecuencia en el análisis multivariante. La decisión de emplear sólo observaciones con datos completos tiene una naturaleza conservadora y «segura», pero como se ha mostrado en las páginas anteriores, existen limitaciones inherentes y sesgos en este método, así como en otras aproximaciones. El investigador no tiene un único método que se ajuste mejor a cada situación, sino que debe hacer un juicio razonado de la situación, considerando todos los factores descritos anteriormente.

Casos atípicos

Los casos atípicos son observaciones con una combinación única de características identificables que les diferencia claramente de las otras observaciones. Los casos atípicos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y deben ser evaluados por los tipos de información que pueden proporcionar. Cuando son benéficos, los casos atípicos, aunque diferentes a la mayor parte de la muestra, pueden ser indicativos de las características segmento de la población que se llegarían a descubrir en el curso normal del análisis. Por el contrario, los casos atípicos problemáticos no son representativos de la población y están en contra de los objetivos del análisis. Los casos atípicos problemáticos pueden distorsionar seriamente los test estadísticos. Debido a la variabilidad en la evaluación de los casos atípicos, se hace imperativo que el investigador examine los datos en busca de la presencia de casos atípicos con el fin de averiguar el tipo de influencia que ejercen. Se remite al lector a las exposiciones del Capítulo 4 y al apéndice de ese capítulo, que aborda el problema de las observaciones influyentes. En estas discusiones, los casos atípicos se sitúan en un escenario particularmente apropiado para evaluar la influencia de las observaciones individuales y determinar si esta influencia es benéfica o dañina.

¿Por qué suceden los casos atípicos? Pueden ser clasificados en cuatro categorías. La primera categoría contiene aquellos casos atípicos que surgen de un error de procedimiento, tales como la entrada de datos o un error de codificación. Estos tipos de casos atípicos deberían identificarse en el nivel de filtrado de datos, pero si se pasan por alto, deberían eliminarse o recodificarse como datos ausentes. La segunda clase de caso atípico es la observación que ocurre como consecuencia de un acontecimiento extraordinario. En este caso, existe una explicación para la unicidad de la observación. El investigador debe decidir si el caso atípico debería ser representado en la muestra. Si es así, el caso atípico tiene que ser retenido en el análisis; si no es así, hay que suprimirlo. La tercera clase de caso atípico comprende las observaciones extraordinarias para las que el investigador no tiene explicación. Aunque estos son los casos atípicos más apropiados para ser omitidos, pueden retenerse si el investigador cree que representan un segmento válido de la población. La cuarta y última clase de caso atípico contiene las observaciones que se sitúan fuera del rango ordinario de valores de cada variable pero que son únicos en su combinación de valores entre las variables. En estas situaciones, el investigador debería retener la observación a menos que se disponga de evidencia específica que excluya al caso atípico como un miembro válido de la población.

Las siguientes secciones detallan los métodos usados en la detección de casos atípicos para las situaciones univariantes, bivariantes y multivariantes. Una vez que los casos atípicos han sido identificados, pueden especificarse para ayudar a situarlos en una de las cuatro clases anteriormente descritas. Finalmente, el investigador debe decidir sobre la retención o exclusión de cada caso atípico, juzgando no sólo las características del caso atípico, sino también los objetivos del análisis.

Detección de casos atípicos

Los casos atípicos pueden identificarse desde una perspectiva univariante, bivariante o multivariante. El investigador debería utilizar cuantas perspectivas sean posibles, buscando una consistencia entre los métodos de identificación de casos atípicos. La siguiente exposición detalla los procesos resultantes de cada una de estas tres perspectivas.

Detección univariante

La perspectiva univariante de identificación de casos atípicos examina la distribución de observaciones, seleccionando como casos atípicos aquellos casos que caigan fuera de los rangos de la

distribución. La cuestión principal consiste en el establecimiento de un umbral para la designación como caso atípico. El enfoque típico convierte en primer lugar los valores de los datos en valores estándar, que tienen una media cero y una desviación estándar de uno.

Dado que los valores están expresados en un formato estandarizado, se pueden realizar fácilmente comparaciones entre las variables. Para muestras pequeñas (de 80 o incluso menos observaciones), las pautas sugeridas identifican como atípicos aquellos casos con valores estándar de 2,5 o superiores. Cuando los tamaños muestrales son mayores, las pautas sugieren que el valor umbral del estandarizado se sitúe entre 3 y 4. Si no se usan los valores estándares, entonces el investigador puede identificar los casos que tienen lugar fuera de las gamas de 2,5 frente a 3 o 4 desviaciones estándares, lo cual depende del tamaño muestral. En cualquier caso, el investigador debe darse cuenta que normalmente puede ocurrir que un cierto número de observaciones caigan fuera de esos rangos de la distribución. El investigador debería esforzarse para identificar sólo aquellas observaciones verdaderamente distintivas y designarlas como casos atípicos.

Detección bivariante

Además de la evaluación univariante, pueden evaluarse conjuntamente pares de variables mediante un gráfico de dispersión. Casos que caigan manifiestamente fuera del rango del resto de las observaciones pueden identificarse como puntos aislados en el gráfico de dispersión. Para ayudar a determinar el rango esperado de las observaciones, se puede superponer sobre el gráfico de dispersión una elipse que represente un intervalo de confianza especificado (variando entre 50 y 90 por ciento de la distribución) para una distribución normal bivariante. Esto proporciona una representación gráfica de los límites de confianza y facilita la identificación de casos atípicos. El gráfico de influencia es otra variante del gráfico de dispersión. En éste, cada punto varía en tamaño según su influencia en las relaciones. Estos métodos proporcionan cierta evaluación de la influencia de cada observación que complementa a la designación de casos como casos atípicos.

Detección multivariante

La tercera perspectiva para identificar casos atípicos implica una evaluación multivariante de cada observación a lo largo de un conjunto de variables. Dado que la mayoría de los análisis multivariantes tienen más de dos variables, el investigador necesita una forma de medición objetiva de la posición multidimensional de cada observación relativa a un punto común. La medida D^2 de Mahalanobis puede usarse con este fin. La D^2 de Mahalanobis es una medida de la distancia de cada observación en un espacio multidimensional respecto del centro medio de las observaciones. Debido a que proporciona una medida común de centralidad multidimensional, también tiene propiedades estadísticas que tienen en cuenta las pruebas de significación. Dada la naturaleza de los test estadísticos, se sugiere que se use un nivel muy conservador, quizá 0,001, como valor umbral para la designación como caso atípico.

Designación como caso atípico

Cuando las observaciones candidatas a la designación como un caso atípico han sido identificadas por métodos univariantes, bivariantes o multivariantes, el investigador debe entonces seleccionar aquellas observaciones que demuestran una unicidad real en comparación con el resto de la población. El investigador debería abstenerse de designar muchas observaciones como casos atípicos y no debería caer en la tentación de eliminar aquellos casos que no son consistentes con los casos restantes, simplemente porque son diferentes.

Descripción de casos atípicos y especificación

Una vez que se han identificado los potenciales casos atípicos, el investigador debería generar identificaciones de cada observación atípica y examinar cuidadosamente que los datos de las variables responsables sean casos atípicos. Además de estos exámenes visuales, el investigador puede emplear también técnicas multivariantes como el análisis discriminante o la regresión múltiple para identificar las diferencias entre los casos atípicos y las otras observaciones. El investigador debería continuar este análisis hasta que sea satisfactorio el aspecto de los datos que distinguen el caso atípico del resto de las observaciones. Si es posible debería asignar el caso atípico a uno de los cuatro tipos ya estudiados.

Mantenimiento o eliminación de los casos atípicos

Una vez que se han identificado, especificado y catalogado los casos atípicos, el investigador debe decidir entre mantenerlos o destruirlos. Hay muchas doctrinas entre los investigadores sobre como tratar con los casos atípicos. Nuestra creencia es que deberían mantenerse a menos que exista una prueba demostrable de que son verdaderas aberraciones y no son representativos de las observaciones de la población. Pero si representan a un segmento de la población, deberían retenerse para asegurar su generalidad al conjunto de la población. Si se eliminan los casos atípicos, el investigador corre el riesgo de mejorar el análisis pero limitar su generalidad. Si los casos atípicos son problemáticos en una técnica particular, muchas veces pueden ser manejados de una forma tal que se ajusten al análisis sin que lo distorsionen significativamente.

Un ejemplo del análisis de casos atípicos

Como ejemplo de la detección de casos atípicos, examinaremos ahora las observaciones de la base de datos de HATCO introducidas en el Capítulo 1 para casos atípicos. Las variables consideradas en el análisis son las variables métricas X_1, X_2, X_3, X_4, X_5 X_6, X_7 y X_9. El análisis de casos atípicos tendrá en cuenta diagnóstico univariante, bivariante y multivariante. Si se encuentran candidatos para la designación como casos atípicos, serán examinados y se procederá a decidir su mantenimiento o eliminación.

Detección univariante y bivariante

El primer paso es examinar las observaciones de cada una de las variables individualmente. La Tabla 2.9 contiene las observaciones con valores estandarizados de variables que exceden $\pm 2{,}5$. Desde esta perspectiva univariante, sólo unas pocas observaciones exceden el umbral en una sola variable, pero ninguna observación fue un caso atípico univariante sobre más de una variable. Desde una perspectiva bivariante, los gráficos de dispersión están formados para X_1, X_2, X_3, X_4, X_5 X_6 y X_7 versus X_9, una de las variables métricas usadas como una variable dependiente en muchas de las técnicas multivariantes. Se superpone una elipse que representa el 90 por ciento del intervalo de confianza de una distribución normal bivariante sobre el gráfico de dispersión (véase Figura 2.6). La segunda parte de la Tabla 2.9 contiene las observaciones que quedan fuera de este elipse. Este es el 90 por ciento de un intervalo de confianza; por tanto esperaríamos que algunas observaciones caigan fuera de la elipse. Sin embargo, varias observaciones (3, 5, 57 y 96) aparecen varias veces, quizá indicando que existen casos atípicos bivariantes.

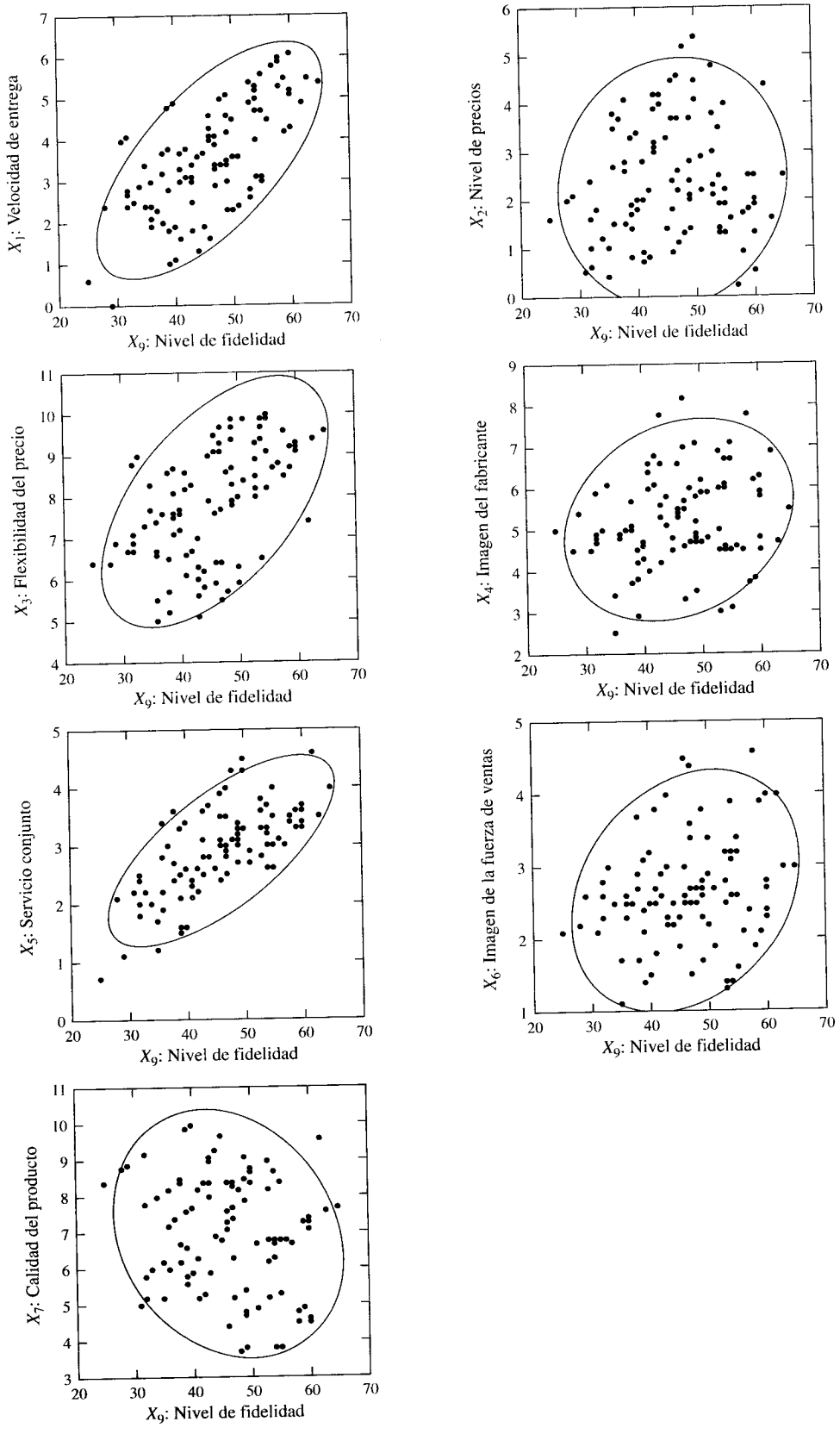


FIGURA 2.6. Identificación gráfica de casos atípicos bivariantes.

TABLA 2.9. Identificación de casos atípicos univariantes y bivariantes

Casos atípicos univariantes con valores estandarizados (valores Z) que exceden ±2,5	
Variable	*Casos*
X_1	39
X_2	71
X_3	ninguno
X_4	82
X_5	96
X_6	5, 42
X_7	ninguno
X_9	ninguno
X_{10}	ninguno

Casos atípicos bivariantes que están fuera de la elipse de intervalo de confianza del 90 por ciento	
X_9 *con*	*Casos*
X_1	1, 39, 95, 96
X_2	3, 49, 57, 7, 96, 97
X_3	11, 57, 96, 100
X_4	5, 22, 42, 50, 72, 82, 93, 96
X_5	3, 22, 39, 57, 71, 96
X_6	5, 7, 42, 82, 96
X_7	57, 58, 95, 96

Detección multivariante

El último método de diagnóstico es evaluar los casos atípicos multivariantes con la medida D^2 de Mahalanobis (véase Tabla 2.10). Este método evalúa la posición de cada observación comparada con el centro de todas las observaciones de un conjunto de variables. En este caso, se usaron para la evaluación todas las variables métricas. Como ya se ha apuntado, los test de significación estadística con esta medida son muy conservadores (excediendo 0,001). Con este umbral, se identifican dos observaciones (22 y 55) como significativamente diferentes. Es interesante resaltar que estas observaciones no fueron apreciadas en el análisis univariante y bivariante y que aparecen sólo en los test multivariantes. Esto indica que no son únicas en cada variable aislada sino que son únicas en la combinación de varibales.

Mantenimiento o eliminación de casos atípicos

Como resultado de estos test de diagnóstico, ninguna observación parece mostrar las características de un caso atípico que debiera ser eliminado. Cada variable tiene algunas observaciones que son extremas, y que deberían considerarse si se va a utilizar la variable en el análisis. Pero ninguna de las observaciones son extremas sobre un número suficiente de variables como para ser consideradas no representativas de la población. En todos los casos, las observaciones denominadas como casos atípicos, incluso con los test multivariantes, parecen suficientemente similares al resto de las variables como para retenerlas en el análisis multivariante. No obstante, el investigador debería siempre examinar los resultados de cada técnica específica para identificar observaciones que pueden llegar a ser atípicas en esa aplicación particular.

Verificación de los supuestos del análisis multivariante

La última etapa del examen de los datos consiste en la comprobación de los supuestos subyacentes básicos del análisis multivariante. La necesidad de comprobar los supuestos estadísticos aumenta en las aplicaciones multivariantes como consecuencia de dos características de este tipo de análisis. En primer lugar, la complejidad de las relaciones, debido al uso habitual de una gran cantidad de variables, hace que las distorsiones y los sesgos potenciales sean más potentes cuando se incumplan los supuestos. En realidad, las violaciones combinadas llegar a ser incluso más perju-

TABLA 2.10. Identificación de casos atípicos multivariantes

Número de casos	*Mahalanobis* D^2	D^2/gl	*gl*	*Significación*	*Número de casos*	*Mahalanobis* D^2	D^2/gl	*gl*	*Significación*
1	7,031	1,004	7	0,4256	51	6,362	0,909	7	0,4982
2	6,691	0,956	7	0,4617	52	8,467	1,210	7	0,2932
3	7,567	1,081	7	0,3723	53	6,913	0,988	7	0,4380
4	7,103	1,015	7	0,4182	54	3,244	0,463	7	0,8615
5	12,870	1,839	7	0,0753	55	35,197	5,028	7	0,0000
6	0,517	0,931	7	0,4809	56	3,082	0,440	7	0,8773
7	8,634	1,233	7	0,2800	57	10,488	1,498	7	0,1626
8	6,563	0,938	7	0,4758	58	5,265	0,752	7	0,6276
9	6,375	0,911	7	0,4967	59	4,348	0,621	7	0,7390
10	3,626	0,518	7	0,8217	60	7,012	1,002	7	0,4276
11	4,237	0,605	7	0,7522	61	13,001	1,857	7	0,0721
12	3,389	0,484	7	0,8468	62	5,798	0,828	7	0,5635
13	3,768	0,538	7	0,8061	63	3,322	0,475	7	0,8537
14	5,030	0,719	7	0,6563	64	6,926	0,989	7	0,4367
15	8,962	1,280	7	0,2554	65	11,683	1,669	7	0,1115
16	6,398	0,914	7	0,4942	66	2,109	0,301	7	0,9536
17	7,212	1,030	7	0,4071	67	4,382	0,626	7	0,7349
18	5,350	0,764	7	0,6173	68	5,925	0,846	7	0,5486
19	5,899	0,843	7	0,5516	69	4,878	0,697	7	0,6749
20	8,962	1,280	7	0,2554	70	5,057	0,722	7	0,6530
21	2,978	0,425	7	0,8870	71	8,294	1,185	7	0,3074
22	35,390	5,056	7	0,0000	72	10,095	1,442	7	0,1833
23	8,333	1,190	7	0,3042	73	5,887	0,841	7	0,5530
24	2,974	0,425	7	0,8874	74	5,363	0,766	7	0,6157
25	4,909	0,701	7	0,6711	75	6,471	0,924	7	0,4859
26	3,463	0,495	7	0,8391	76	4,925	0,704	7	0,6691
27	3,171	0,453	7	0,8687	77	5,847	0,835	7	0,5577
28	5,765	0,824	7	0,5674	78	7,522	1,075	7	0,3766
29	7,601	1,086	7	0,3691	79	12,279	1,754	7	0,0918
30	5,188	0,741	7	0,6370	80	2,270	0,324	7	0,9434
31	2,751	0,393	7	0,9071	81	4,943	0,706	7	0,6669
32	7,024	1,003	7	0,4264	82	14,118	2,017	7	0,0491
33	5,678	0,811	7	0,5778	83	6,837	0,977	7	0,4460
34	3,529	0,504	7	0,8321	84	2,366	0,338	7	0,9369
35	6,539	0,934	7	0,4784	85	3,016	0,431	7	0,8835
36	2,900	0,414	7	0,8941	86	3,493	0,499	7	0,8359
37	6,704	0,958	7	0,4603	87	3,354	0,479	7	0,8504
38	3,030	0,433	7	0,8823	88	2,417	0,345	7	0,9332
39	10,213	1,459	7	0,1768	89	6,011	0,859	7	0,5385
40	3,827	0,547	7	0,7995	90	4,860	0,694	7	0,6771
41	2,898	0,414	7	0,8943	91	3,763	0,538	7	0,8067
42	12,282	1,755	7	0,0917	92	5,841	0,834	7	0,5584
43	7,129	1,018	7	0,4156	93	14,328	2,047	7	0,0456
44	4,819	0,688	7	0,6821	94	5,407	0,772	7	0,6105
45	6,670	0,953	7	0,4640	95	7,391	1,056	7	0,3893
46	7,475	1,068	7	0,3811	96	16,708	2,387	7	0,0194
47	14,094	2,013	7	0,0495	97	8,195	1,171	7	0,3157
48	6,152	0,879	7	0,5221	98	4,990	0,713	7	0,6612
49	7,561	1,080	7	0,3729	99	5,587	0,798	7	0,5888
50	9,029	1,290	7	0,2506	100	4,704	0,672	7	0,6960

gl = Grados de libertad.
Mahalanobis D^2 valor que se basa en las siguientes variables (X_1, X_2, X_3, X_4, X_5, X_6 y X_7). El valor D^2/gl está distribuido aproximadamente como un valor *t*.

diciales si se consideran separadamente. En segundo lugar, la complejidad de los análisis y de los resultados pueden enmascarar los «signos» de las violaciones de los supuestos que son aparentes en los más sencillos análisis multivariantes. En casi todos los ejemplos, los procedimientos multivariantes estimarán el modelo multivariante y producirán resultados incluso cuando los supuestos se vean severamente incumplidos. Por tanto, el investigador debe estar atento a cualquier incumplimiento de los supuestos y a las implicaciones que puedan tener para el proceso de estimación o interpretación de los resultados.

Valoración de las variables individuales frente al modelo univariante

El análisis multivariante requiere que los supuestos subyacentes a las técnicas estadísticas sean contrastados dos veces: en primer lugar para las variables aisladas, semejante a las pruebas de los supuestos del análisis univariante, y en segundo lugar para el **valor teórico** del modelo multivariante, que actúa colectivamente sobre las variables a analizar y por tanto debe cumplir los mismos supuestos que las variables individuales. Este capítulo se centra en el examen de las variables univariantes en relación al cumplimiento de los supuestos subyacentes a los procedimientos multivariantes. En cada capítulo se analizarán los métodos utilizados para evaluar los supuestos subyacentes al valor teórico de cada técnica multivariante.

Normalidad

El supuesto fundamental del análisis multivariante es la **normalidad** de los datos, en referencia al perfil de la distribución de los datos para una única variable métrica y su correspondencia con una **distribución normal**, punto de referencia de los métodos estadísticos. Si la variación respecto de la distribución normal es suficientemente amplia, todos los test estadísticos resultantes no son válidos, dado que se requiere la normalidad para el uso de los estadísticos de la t y de la F. Tanto los métodos estadísticos univariantes como los multivariantes analizados en este texto se basan en el supuesto de la normalidad univariante, suponiendo también los multivariantes la normalidad multivariante. La normalidad univariante para una única variable es fácil de contrastar, siendo posible varias medidas correctoras, tal y como se demuestra más adelante. Dicho de forma sencilla, la normalidad multivariante (la combinación de dos o más variables) implica que las variables individuales son normales en un sentido univariante y que sus combinaciones también sean normales. Por tanto, si una variable es una normal multivariante, es también normal univariante. Sin embargo, lo contrario no es necesariamente cierto (dos o más variables normales univariantes no son necesariamente normal multivariante). Por tanto, una situación en la que todas las variables exhiben normalidad univariante ayudará a obtener normalidad multivariante, aunque no la garantiza. La normalidad multivariante es mucho más difícil de contrastar, aunque existen varios test para situaciones en las que la técnica multivariante se ve particularmente afectada por una violación de los supuestos. En este texto, nos centraremos en evaluar y alcanzar la normalidad univariante para todas las variables y acudiremos a la multivariante cuando sea especialmente crítica. Incluso aunque las muestras grandes tiendan a disminuir los efectos perniciosos de la no normalidad, el investigador debería evaluar la normalidad de todas las variables incluidas en el análisis.

Análisis gráfico de la normalidad

El test más simple para diagnosticar la normalidad es una comprobación visual del histograma que compare los valores de los datos observados con una distribución aproximada a la distribución normal (véase Figura 2.1). Aunque atractivo por su simplicidad, este método es problemático para muestras pequeñas, donde la construcción del histograma (por ejemplo, el número de categorías

o la anchura de las categorías) puede distorsionar la representación visual de tal forma que el análisis sea inútil. Una aproximación de mayor confianza es el **gráfico de distribución normal**, que compara la distribución acumulada de los valores reales de los datos con la distribución acumulada de una distribución normal. La distribución normal sigue una línea recta en diagonal, comparándola con el gráfico de los valores de los datos. Si una distribución es normal, la línea que representa la distribución real de los datos sigue de cerca a la diagonal.

La Figura 2.7 muestra varios gráficos de distribución normal y la correspondiente distribución univariante de la variable. En los gráficos de distribución normal se representa una característica del perfil de la distribución, la curtosis. La **curtosis** se refiere al «apuntamiento» o «llanura» de la distribución, comparada con la distribución normal. Cuando la línea cae por debajo de la diagonal, la distribución es más llana de lo esperado (platicúrtica). Cuando la línea cae por encima de la diagonal, la distribución es más puntiaguda que la curva normal (leptocúrtica). Por ejemplo, en el gráfico de distribución normal de una distribución normal (Figura 2.7d), vemos una curva con un nítido perfil en S. Inicialmente la distribución es más plana, y la línea cae por debajo de la diagonal. Entonces, la parte puntiaguda de la distribución se mueve rápidamente por encima de la diagonal y se desplaza otra vez por debajo de la diagonal a medida que la distribución se aplana. Una distribución no puntiaguda (platicúrtica) tiene una pauta opuesta (Figura 2.7c). Otro modelo común es un simple arco, tanto por encima como por debajo de la diagonal, que indica la simetría de la distribución. Una simetría negativa (Figura 2.7e) se indica mediante un arco por debajo de

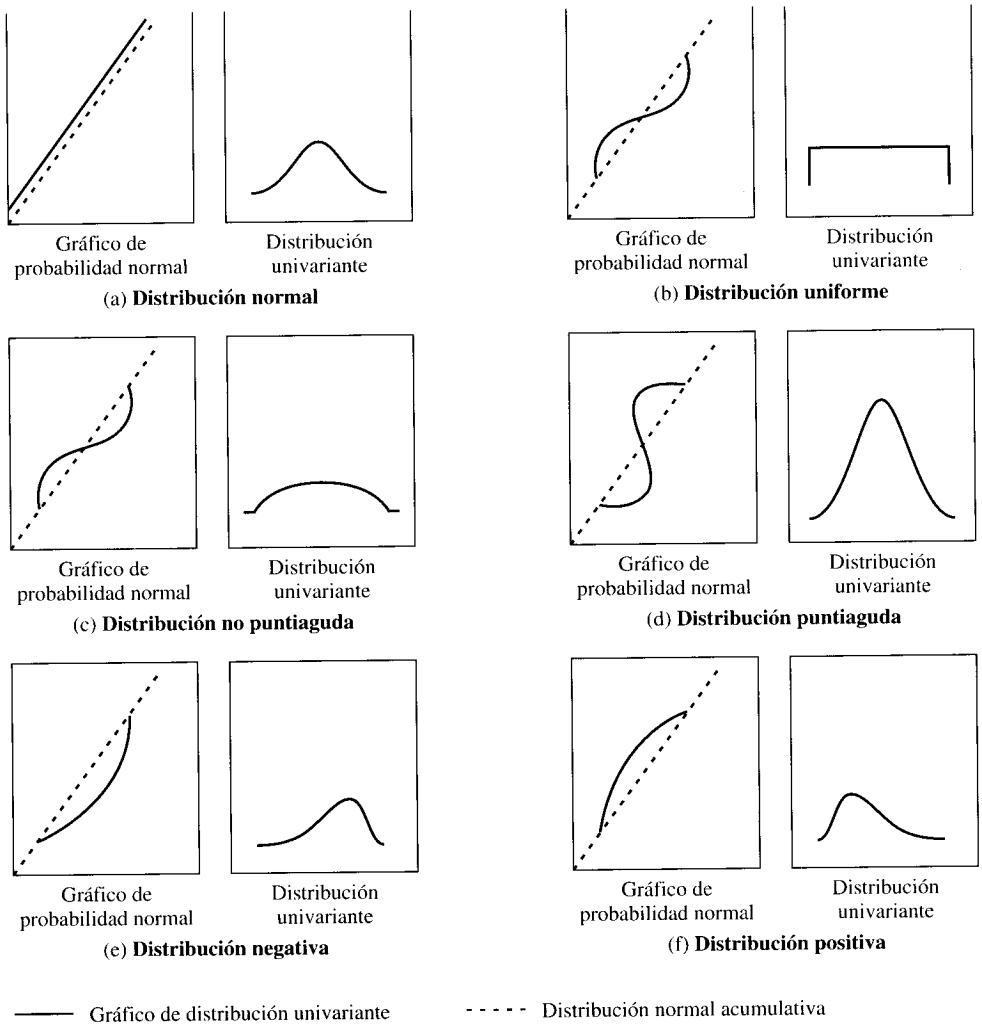


FIGURA 2.7. Gráficos de distribución normal y las correspondientes distribuciones univariantes.

la diagonal, mientras que un arco por encima de la diagonal representa una distribución positivamente simétrica (Figura 2.7f). Una excelente fuente para interpretar los gráficos de distribución normal que muestren los diversos modelos e interpretaciones es Daniel y Wood [5]. Estos modelos específicos no sólo identifican la no normalidad, sino que también nos dicen la forma de la distribución original y la solución apropiada a aplicar.

Test estadísticos de normalidad

Además de examinar el gráfico de distribución normal, pueden utilizarse también test estadísticos para evaluar la normalidad. El test más simple es una regla basada en el valor de simetría (disponible como parte de los estadísticos descriptivos básicos para una variable procesada en todos los programas estadísticos). El valor estadístico (z) se calcula como

$$z_{\text{simetria}} = \frac{\text{simetría}}{\sqrt{6/N}}$$

donde N es el tamaño de la muestra. Un valor z también puede ser calculado para el valor de curtosis utilizando la siguiente formula:

$$z_{\text{curtosis}} = \frac{\text{curtosis}}{\sqrt{24/N}}$$

Si el valor calculado de z excede un valor crítico, entonces la distribución es no normal por lo que se refiere a esta característica. El valor crítico es de una distribución z, basada en los niveles de significación que deseemos. Por ejemplo, un valor calculado que exceda $\pm$ 2,58 indica que podemos rechazar el supuesto sobre la normalidad de la distribución a un nivel de probabilidad de 0,01. Otro valor crítico habitualmente utilizado es $\pm$ 1,96, que corresponde a un nivel de error de 0,05.

Los test estadísticos específicos se encuentra también en SPSS, SAS, BMDP y la mayor parte de los programas. Los dos más comunes son el test de Shapiro-Wilks y una modificación del test de Kolmogorov-Smirnov. Cada uno calcula el nivel de significación para las diferencias respecto a una distribución normal. El investigador debería siempre recordar que los test de significación son menos útiles en muestras pequeñas (menores de 30) y muy sensibles para grandes muestras (superiores a 1.000 observaciones). Por tanto, el investigador debería siempre usar tanto los gráficos como cualquier comprobación estadística para evaluar el grado real de desviación de la normalidad.

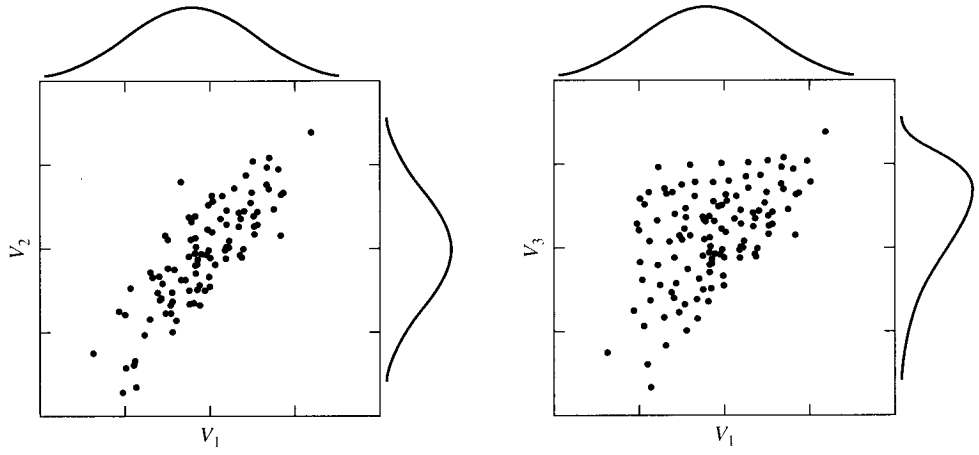
Soluciones para la no normalidad

Existen algunas transformaciones de datos para acomodar las distribuciones no normales que se expondrán posteriormente. Este capítulo se limita a la discusión de los test de normalidad univariante y las transformaciones. Sin embargo, cuando examinemos los otros métodos multivariantes, tales como la regresión múltiple o el análisis multivariante de la varianza, discutiremos también los test de normalidad multivariante. Más aún, en muchas de las ocasiones en que se indica no normalidad es en realidad un resultado de otras violaciones de los supuestos; por tanto, remediando los otros incumplimientos eliminamos el problema de la normalidad. Por esta razón, debería realizar test de normalidad después o junto con los análisis y soluciones para las otras violaciones. (Para aquellos interesados en la normalidad multivariante, véase [8, 11].)

Homocedasticidad

La **homocedasticidad** es un supuesto relativo primordialmente a las relaciones de dependencia entre variables. Se refiere al supuesto de que las variables dependientes exhiban iguales niveles de varianza a lo largo del rango del predictor de la variables. La homocedasticidad es deseable porque la varianza de la variable dependiente que se está explicando en la relación de dependencia no debería concentrarse sólo en un limitado rango de los valores independientes. Aunque las variables dependientes deben ser métricas, este concepto de igual dispersión de la varianza a lo largo de las variables independientes puede aplicarse cuando las variables son métricas o no métricas. Con variables independientes métricas, el concepto de homocedasticidad se basa en la dispersión de la varianza de la variable dependiente a lo largo del rango de los valores de la variable independiente, que se encuentra en técnicas como la regresión múltiple. El mismo concepto se aplica también cuando las variables independientes son no métricas. En estos casos, tal y como se encuentran en ANOVA y MANOVA, el centro es ahora la igualdad de la varianza (una variable dependiente) o la matriz de varianza/covarianza (varias variables independientes) a lo largo de los grupos formados por las variables independientes no métricas. La igualdad de las matrices de varianza/covarianza se observa también en el análisis discriminante, pero en esta técnica el énfasis es en la dispersión de las variables independientes a lo largo de los grupos formados por la medida dependiente no métrica. En cada uno de estos casos, el propósito es el mismo: asegurar que la varianza usada en la explicación y predicción se disperse a través del rango de valores, permitiendo así un «test limpio» de las relaciones a lo largo de todos los valores de las variables no métricas.

En la mayoría de las situaciones, tenemos diferentes valores de la variable dependiente para cada valor de la variable independiente. Para que esta relación se capte completamente, la dispersión (varianza) de los valores de la variable dependiente debe ser igual para cada valor de la variable predictor. La mayoría de los problemas con varianzas desiguales surgen de una de estas dos fuentes. La primera es el tipo de variables incluidas en el modelo. Por ejemplo, a medida que una variable aumenta en valor (es decir, cuando las unidades van desde cero a millones), existe un rango más amplio de respuestas posibles para los valores más elevados. La segunda fuente surge de una distribución simétrica que crea heterocedasticidad. En la Figura 2.8a, los gráficos de dispersión de puntos de los datos para dos variables (V_1 y V_2) con distribuciones normales exhiben la misma dispersión a lo largo de todos los valores de los datos (es decir, homocedasticidad). Sin embargo, en la Figura 2.8b, observamos también una dispersión desigual (**heterocedasticidad**) provocada por la simetría de una de las variables (V_2). Para diferentes valores de V_1, tenemos diferentes pautas de dispersión para V_2. Esto provocará que las predicciones sean mejores a ciertos niveles de la variable independiente que a otros. Violando este supuesto a menudo realizamos unos test de las hipótesis muy conservadores o demasiado sensibles.



(a) Homocedasticidad (b) Heterocedasticidad

FIGURA 2.8. Gráficos de dispersión de relaciones de homocedasticidad y heterocedasticidad.

El efecto de la heterocedasticidad está a menudo también relacionado con el tamaño de la muestra, especialmente cuando examinamos la dispersión de la varianza entre grupos. Por ejemplo, en ANOVA o MANOVA, el impacto de la heterocedasticidad de los test estadísticos depende de los tamaños de la muestra asociados con los grupos de menor o mayor varianza. En el análisis de la regresión múltiple ocurrirán efectos similares en distribuciones altamente simétricas donde existan un número desproporcionado de encuestados en ciertos rangos de la variable independiente.

Test gráficos de igual dispersión de la varianza

La prueba de homocedasticidad de dos variables métricas se evalúa mejor gráficamente. La aplicación más común de esta forma de evaluación se produce en la regresión múltiple, en relación con la dispersión de la variable dependiente a lo largo de las variables independientes métricas. Dado que el eje del análisis de la regresión es el valor teórico, el gráfico de residuos se usa para revelar la presencia de homocedasticidad (o su opuesto, heterocedasticidad, desigual dispersión de la varianza). En el Capítulo 4 se detallan estos procedimientos en la discusión del análisis de los residuos. Los gráficos de cajas sirven bien para representar el grado de variación entre los grupos formados por una variable categórica. El largo de la caja y de los bigotes indica la variación de los datos entre este grupo.

Test estadísticos de homocedasticidad

Los test estadísticos de igual dispersión de la varianza se refieren a la varianza en grupos formados por variables métricas. El test más común, el test de Levene, puede usarse para evaluar si las varianzas de una única variable métrica son iguales a lo largo de cualquier cantidad de grupos. Si se está contrastando más de una variable métrica, implicando la comparación de la igualdad de las matrices de varianzas/covarianzas, se aplica el test *M* de Box. El test *M* de Box existe tanto en el análisis multivariante como en el análisis discriminante y se discute con más detalle en los capítulos dedicados a estas técnicas.

Soluciones para la heterocedasticidad

Los problemas de heterocedasticidad pueden solucionarse a través de transformaciones de datos, similares a las usadas para conseguir la normalidad. Como ya se ha mencionado, en muchas ocasiones la heterocedasticidad es el resultado de la no normalidad de una de las variables, y la corrección de la no normalidad resuelve igualmente la dispersión de la varianza. En una sección posterior se discuten las transformaciones de los datos de las variables que hacen que todos los valores tengan un efecto potencialmente igual en la predicción.

Linealidad

La **linealidad** es un supuesto implícito de todas las técnicas multivariantes basadas en medidas de correlación, incluyendo la regresión múltiple, regresión logística, análisis factorial y los modelos de ecuaciones estructurales. Dado que las correlaciones representan sólo la asociación lineal entre variables, los efectos no lineales no estarán representados en el valor de la correlación. Como resultado, es siempre prudente examinar todas las relaciones para identificar cualquier desplazamiento de la linealidad que pueda impactar la correlación.

Identificación de relaciones no lineales

La forma más común de evaluar la linealidad es examinar los gráficos de dispersión de las variables e identificar cualquier pauta no lineal en los datos. Una aproximación alternativa es ir a un

análisis de regresión múltiple (los detalles de esta técnica se ven en el Capítulo 4) y examinar los residuos. Los residuos reflejan la parte no explicada de la variable dependiente; por tanto, cualquier parte no lineal de la relación quedará reflejada en los residuos. El examen de los residuos puede aplicarse a la regresión múltiple, donde el investigador puede detectar cualquier efecto no lineal no representado en el valor teórico de la regresión. Una discusión más detallada del análisis de residuos está en el Capítulo 4.

Soluciones para la no linealidad

Si se detecta una relación no lineal, la aproximación más directa es transformar una o ambas variables para conseguir la linealidad. Posteriormente en este capítulo, se discutirán unas cuantas transformaciones. Una alternativa a la transformación de los datos es la creación de una nueva variable que represente la parte no lineal de la relación. El proceso de crear e interpretar estos resultados adicionales, que pueden usarse en todas las relaciones lineales, se discute en el Capítulo 4.

Ausencia de errores correlacionados

La predicciones con cualquiera de las técnicas de dependencia no son perfectas, y rara vez encontraremos una situación donde lo sean. Sin embargo, intentamos asegurar que cualquiera de los errores de predicción no esté correlacionado con el resto. Por ejemplo, si encontramos un indicio que sugiera que los errores son positivos y negativos alternativamente, entenderíamos que hay alguna relación sistemática no explicada de la variable dependiente. Si existe tal situación, no podemos estar seguros de que nuestros errores de predicción sean independientes de los niveles que estamos intentando predecir. Existe otro factor que está afectando los resultados, pero que no está incluido en el análisis.

La identificación de errores correlacionados

Las situaciones más comunes de incumplimiento de este supuesto se deben al proceso de recogida de datos. Factores análogos pueden afectar a un grupo y no afectar a otro. Si se analizan separadamente, los efectos conjuntos son constantes y no influyen en la estimación de la relación. Pero si se combinan las observaciones de ambos grupos, entonces la relación final estimada debe ser un «compromiso» entre los dos tipos de relaciones. Esto provoca un sesgo de los datos porque una causa sin especificar está influyendo en la estimación de la relación.

Para identificar los errores correlacionados, en primer lugar el investigador tiene que identificar las causas posibles. En nuestro ejemplo anterior, la causa vendría de que hay dos grupos separados en la recogida de datos. Una vez que la causa potencial haya sido identificada, el investigador podría observar si existen diferencias entre los grupos. El hallazgo de diferencias en los errores de predicción para los dos grupos sería entonces la base para determinar que un efecto no especificado es la «causa» de los errores correlacionados.

Soluciones para los errores correlacionados

Los errores correlacionados tienen que ser corregidos con la inclusión del factor causante omitido en el análisis multivariante. En nuestro ejemplo anterior, el investigador añadiría una variable para indicar la clase donde estaban los encuestados. La solución más común es la adición de una(s) variable(s) al análisis que representa el factor omitido. La tarea clave a la que se enfrenta el investigador no es la solución en sí, sino la identificación del efecto no especificado y una manera de representarlo en el análisis.

Transformaciones de los datos

Las **transformaciones de los datos** proporcionan un medio de modificar variables por una o dos razones: corregir el incumplimiento de los supuestos estadísticos subyacentes a las técnicas multivariantes o mejorar la relación (correlación) entre variables. La transformación de los datos puede basarse en razones tanto «teóricas» (transformaciones cuya conveniencia se basa en la naturaleza de los datos) como «derivadas de los datos» (donde las transformaciones se sugieren a partir de un examen de los datos). Así, en cada caso el investigador debe proceder muchas veces por ensayo y error, ponderando la mejora frente a la necesidad de transformaciones adicionales.

Todas las transformaciones descritas pueden llevarse a cabo fácilmente mediante simples comandos de todos los programas estadísticos. Nosotros nos centraremos en transformaciones que puedan calcularse de esta forma, aunque también existen métodos más complicados y sofisticados (por ejemplo, véase Box and Cox [2].

Transformaciones de los datos para conseguir la normalidad y la homocedasticidad

Las transformaciones de los datos proporcionan el medio principal de corregir la no normalidad y heterocedasticidad. En ambos casos, la forma de las variables sugiere transformaciones específicas. Para las distribuciones no normales, las dos formas más comunes son las distribuciones «planas» y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, $1/Y$ o $1/X$). Las distribuciones asimétricas pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Normalmente, las distribuciones negativamente simétricas se transforman de forma más efectiva empleando la raíz cuadrada, mientras que por lo general, el logaritmo funciona mejor para la simetría positiva. De todos modos, el investigador debería aplicar todas las transformaciones posibles y luego seleccionar la variable transformada más apropiada.

La heterocedasticidad es un problema asociado a la normalidad, y en muchas casos la solución del problema tiene que ver también con los problemas de normalidad. La heterocedasticidad se debe también a la distribución de la variable(s). Cuando se examinan los residuos del análisis de la regresión buscando la heterocedasticidad, se observa que un indicio de varianzas desiguales es una distribución con perfil de cono de los residuos (véase Capítulo 4 para detalles más específicos del análisis gráfico de los residuos). Si el cono se abre a la derecha, cogemos la inversa; si se abre a la izquierda, cogemos la raíz cuadrada. Algunas transformaciones pueden asociarse con ciertos tipos de datos. Por ejemplo, el recuento de frecuencias sugiere una transformación de raíz cuadrada; las proporciones se transforman mejor por la transformación del arcoseno ($X_{\text{nueva}} = 2$ arcoseno $\sqrt{X_{\text{antiguo}}}$); y un cambio proporcional se maneja mejor tomando el logaritmo de la variable. En todos los casos, una vez que se han realizado las transformaciones, los datos transformados deberían ser contrastados para ver si se ha logrado la solución deseada.

Transformaciones para conseguir la linealidad

Existen numerosos procedimientos para conseguir la linealidad entre dos variables, pero las relaciones no lineales más simples pueden clasificarse en cuatro categorías (véase Figura 2.9). En cada cuadrante, se muestran las transformaciones potenciales para variable dependiente e independiente. Por ejemplo, si las relaciones locales son como las de la Figura 2.9a, se aplica la raíz cuadrada para conseguir la linealidad. Cuando se muestran las posibilidades de transformación múltiple, se empieza con el método más adecuado para cada cuadrante para después bajar hasta que se consigue la linealidad. Una aproximación alternativa consiste en utilizar variables adicionales, denominadas polinómicas, que representan los componentes no lineales. Este método se discute con mayor detalle en el Capítulo 4.

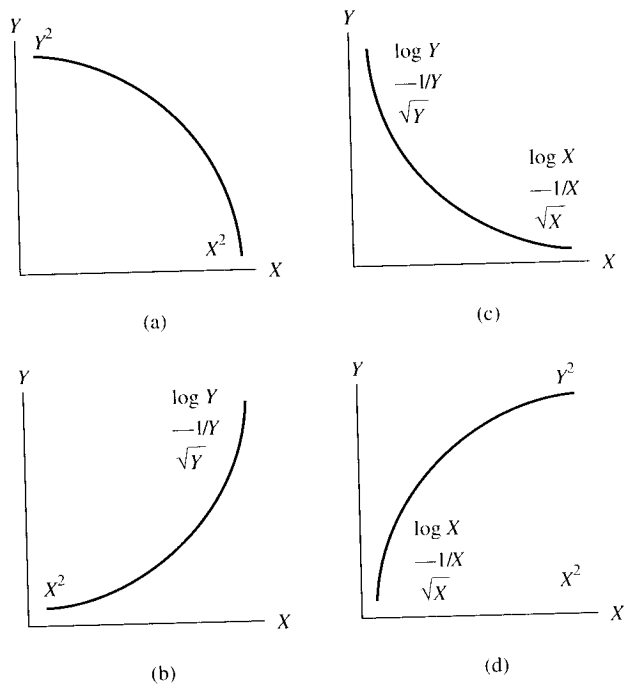


FIGURA 2.9. Selección de transformaciones para conseguir la linealidad.
Fuente: F. Mosteller y J. W. Tukey, *Data Analysis and Regresion*. Reading, Mass.: Addison-Wesley, 1977.

Normas generales para las transformaciones

Existen varios puntos a recordar cuando se llevan a cabo transformaciones de los datos. Incluyen los siguientes:

1. Para obtener un efecto perceptible de la transformación, el ratio entre la medida de la variable y su desviación estándar debería ser menor de 4,0.
2. Cuando la transformación puede realizarse sobre una de las dos variables, seleccionar la variable con el ratio más pequeño del ítem 1.
3. Las transformaciones deberían aplicarse a las variables independientes excepto en el caso de la heterocedasticidad.
4. La heterocedasticidad sólo puede solucionarse mediante la transformación de la variable dependiente en una relación de dependencia. Si una relación heterocedástica es además no lineal, deberían transformarse la variable dependiente y quizá la independiente.
5. Las transformaciones pueden cambiar la interpretación de las variables. Por ejemplo, las variables transformadas tomando sus logaritmos trasladan la relación en una medida de cambio proporcional (elasticidad). Siempre hay que asegurarse la exploración de todas las posibles interpretaciones de las variables transformadas.

Un ejemplo de contrastación de los supuestos básicos del análisis multivariante

Con el fin de ilustrar las técnicas de contrastación de datos para conseguir el cumplimiento de los supuestos subyacentes al análisis multivariante y proporcionar un fundamento en el uso de los datos en los capítulos siguientes, examinaremos el conjunto de datos introducidos en el Capítulo 1. En el curso del análisis, se cubrirán los supuestos de normalidad, homocedasticidad y linealidad.

El cuarto supuesto básico, la ausencia de correlación entre los errores, sólo puede apreciarse en el contexto de un modelo multivariante específico y por tanto, será cubierto en los últimos capítulos para cada técnica multivariante. Se pondrá mayor énfasis en las variables métricas, aunque las variables no métricas serán evaluadas cuando sea apropiado.

Normalidad

El primer análisis a realizar para evaluar la normalidad de las variables métricas es la derivación de los gráficos de probabilidad normal. La Figura 2.10 contiene los gráficos de cada una de las nueve variables. En nuestro examen de los gráficos, observamos ciertos desplazamiento de la diagonal, indicativo de una ausencia de normalidad. En referencia a los modelos vistos en la Figura 2.7, vemos que X_2 parece simétrica positiva, X_3 se aproxima a una distribución uniforme y X_5 parece simétrica negativa.

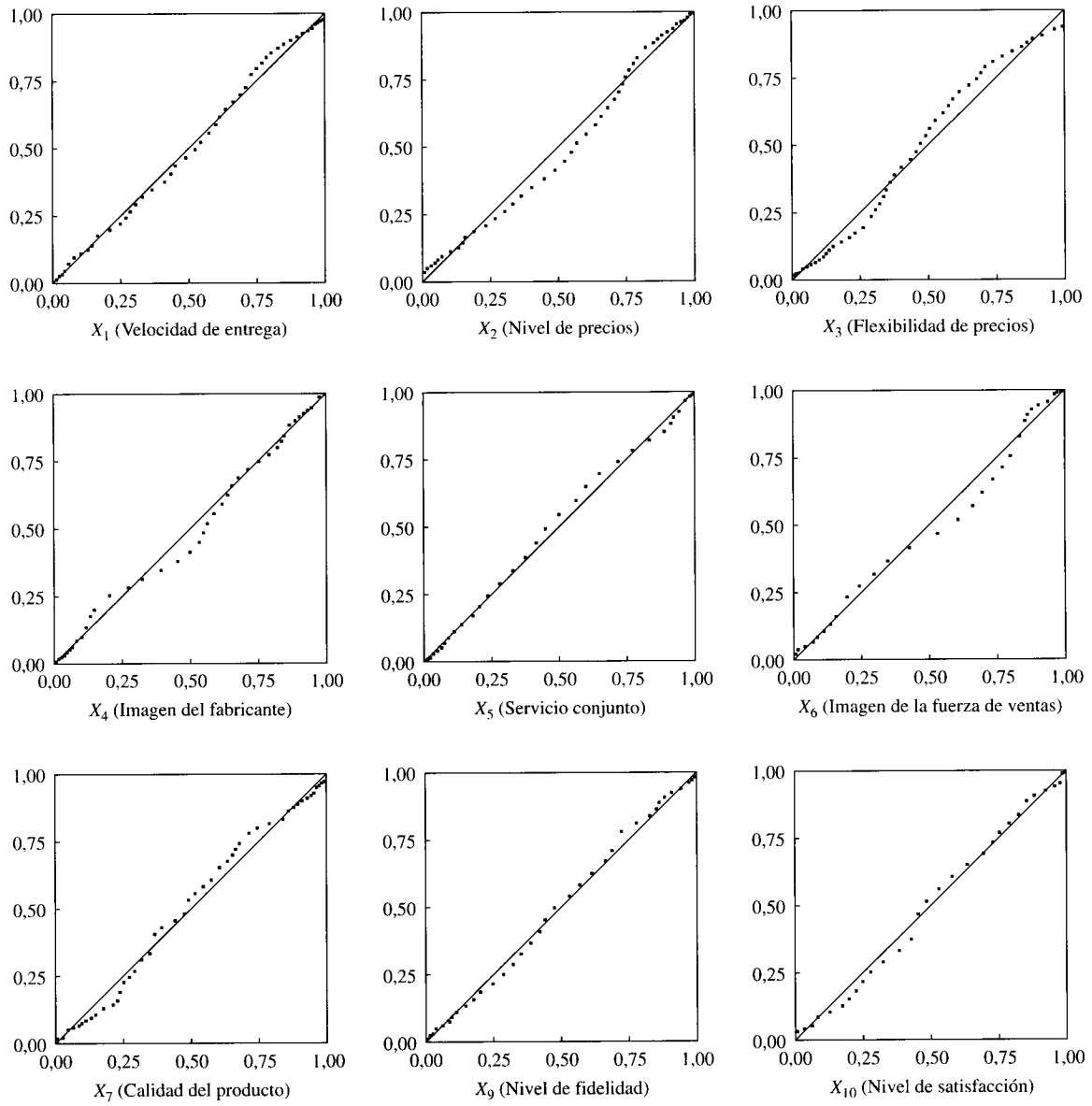


FIGURA 2.10. Gráficos de probabilidad normal de variables métricas.

TABLA 2.11. Características de la distribución, la prueba de la normalidad y posibles soluciones

	Descripción de forma[a]							*Soluciones posibles*	
	Simetría		*Curtosis*		*Test de normalidad*				
Variable	*Estadística*	*Valor z*	*Estadística*	*Valor z*	*Estadística*	*Significación*	*Descripción de la distribución*	*Transformación*	*Significación después de solución*
X_1 Velocidad de entrega	–0,085	–0,35	–0,511	–1,07	0,063	> 0,200	Distribución normal	Ninguna	
X_2 Nivel de precios	0,469	1,95*	–0,509	1,06	0,095	0,028	Simetría positiva	Raíz cuadrada	> 0,200
X_3 Flexibilidad de precios	–0,289	1,19	–1,073	2,24*	0,095	0,027	Acercamiento a la distribución normal	Ninguna	
X_4 Imagen del productor	0,218	0,91	0,085	0,18	0,107	0,007	Simetría positiva ligera	Sin mejora posible	
X_5 Servicio conjunto	–0,373	1,55	0,141	0,29	0,085	0,069	Distribución normal	Ninguna	
X_6 Imagen de la fuerza de ventas	0,493	2,04*	0,107	0,22	0,122	0,001	Cola cargada con simetría positiva	Raíz cuadrada	0,032
X_7 Calidad del producto	–0,229	0,95	–0,850	1,77	0,091	0,041	Ligeramente plana	Ninguna	
X_9 Nivel de fidelidad	–0,069	0,26	–0,725	1,52	0,079	0,131	Distribución normal	Ninguna	
X_{10} Nivel de satisfacción	0,089	0,37	–0,763	1,60	0,078	0,142	Distribución normal	Ninguna	

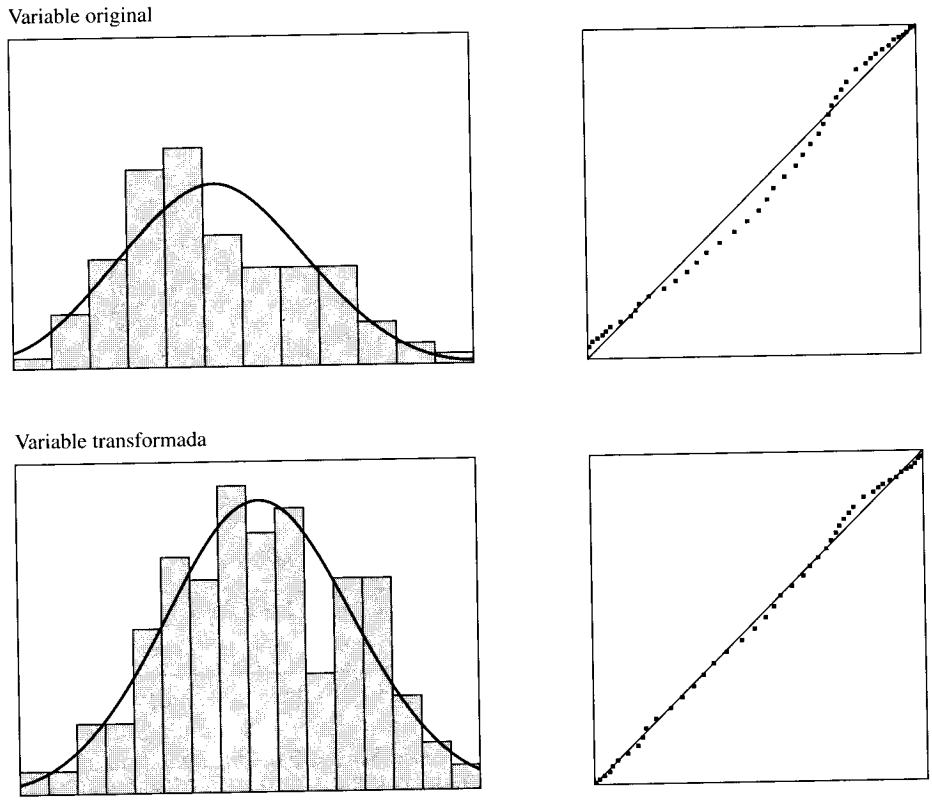
[a] Los valores *z* se derivan con la división de las estadísticas por los errores estándares apropiados de 0,241 (simetría) y 0,478 (curtosis). Se proporcionan las ecuaciones para calcular los errores estándares en el texto.

* Significante al nivel de 0,05.

Podemos complementar este análisis visual con estadísticos que reflejen el perfil de la distribución (simetría y curtosis) así como un test estadístico de normalidad (el test modificado de Kolmogorov-Smirnov). La Tabla 2.11 muestra estos valores para todas las variables métricas. Cuatro variables exhiben un alejamiento estadísticamente significativo de la normalidad (X_2, X_3, X_4 y X_6). La Tabla 2.11 también sugiere el criterio apropiado. Dos variables (X_2 y X_6) fueron transformadas usando la raíz cuadrada. En cada caso, las variables transformadas demuestran la normalidad (véase Tabla 2.11).

La Figura 2.11 demuestra el efecto de la transformación sobre X_2 para alcanzar la normalidad. X_2 aparece notablemente más normal en ambas gráficas, y los descriptores estadísticos también han mejorado. El investigador siempre debe examinar las variables transformadas con igual rigor que las variables originales por lo que se refiere a su normalidad y su forma de distribución.

En el caso de las dos variables restantes (X_3 y X_4), ninguna de las transformaciones podrían mejorar su normalidad. Estas variables tendrán que ser utilizadas en su forma original.



Características de distribución antes y después de la transformación

Forma variable	Descriptores de distribución[a]				Test de normalidad	
	Simetría		Curtosis			
	Estadística	Valor z	Estadística	Valor z	Estadística	Significación
Original X_2	0,469	1,95	–0,509	1,06	0,095	0,028
Transformada X_2	–0,106	0,44	–0,465	0,97	0,062	> 0,200

[a] Los valores z se derivan con la división de las estadísticas por los errores estándares apropiados de 0,241 (simetría) y 0,478 (curtosis). Se proporcionan las ecuaciones para calcular los errores estándares en el texto.

FIGURA 2.11. Transformación de X_2 (nivel de precio) para lograr la normalidad.

En las situaciones en donde la normalidad de estas variables son críticas, los valores transformados pueden usarse con la garantía de que mantienen el supuesto de normalidad. Pero el alejamiento de la normalidad no es tan extremo como para que las variables originales no puedan utilizarse en el análisis en su forma original. Si la técnica es robusta en lo que se refiere a alejamientos de la normalidad, entonces pueden preferirse las variables originales para poder realizar una comparación en la fase de interpretación.

Homocedasticidad

Todos los paquetes estadísticos contienen test para evaluar la homocedasticidad sobre una base univariante (por ejemplo, el test de Levene en SPSS) donde se compara la varianza de una variable métrica a lo largo de los niveles de las variables no métricas. De acuerdo con nuestros objetivos, examinaremos cada una de las variables métricas a lo largo de las cinco variables no métricas en el conjunto de datos. Estos análisis son apropiados en preparación tanto del análisis de la varianza como del análisis multivariante de la varianza donde las variables no métricas son las variables independientes, o el análisis discriminante donde las variables no métricas son las medidas dependientes.

La Tabla 2.12 contiene los resultados del test de Levene para cada una de las variables. Las variables no métricas X_8 y X_{11} muestran ambas heterocedasticidades significativas sobre los mismos factores de desarrollo (X_4, X_5, X_6 y X_7), mientras que X_{12} y X_{14} tienen pocas ocurrencias entre el conjunto completo de variables. Las implicaciones de estos ejemplos de heterocedasticidad deben ser examinadas dondequiera que sean evaluadas las diferencias de grupos, usando tanto variables no métricas como variables independientes, como variables métricas que se consideren dependientes. Si se encontrara algún incumplimiento de los supuestos, existen transformaciones de la variable para ayudar a remediar la dispersión de la varianza.

TABLA 2.12. Test de homocedasticidad

	Variable no métrica/categórica									
	X_8 *Tamaño de la empresa*		X_{11} *Compra de especificación*		X_{12} *Estructura de obtención*		X_{13} *Tipo de industria*		X_{14} *Tipo de situación de compra*	
Variable métrica	*Estadística Levene*	*Sig.*	*Estadística Levene*	*Sig.*	*Estadística Levene*	*Sig.*	*Estadística Levene*	*Sig.*	*Estadística Levene*	*Sig.*
X_1 Velocidad de entrega	0,934	0,336	0,934	0,336	0,382	0,538	0,377	0,540	0,114	0,892
X_2 Nivel de precios	1,582	0,211	1,582	0,211	13,761	0,000	1,345	0,249	8,081	0,001
X_3 Flexibilidad de precios	1,194	0,277	1,194	0,277	4,765	0,031	0,192	0,662	14,383	0,000
X_4 Imagen del productor	6,549	0,012	6,549	0,012	0,281	0,597	0,040	0,842	2,030	0,137
X_5 Servicio conjunto	7,819	0,006	7,819	0,006	5,141	0,026	0,003	0,957	2,888	0,060
X_6 Imagen de la fuerza de ventas	5,279	0,024	5,279	0,024	1,626	0,205	0,264	0,609	1,735	0,182
X_7 Calidad de producto	8,748	0,004	8,748	0,004	4,129	0,045	2,532	0,115	2,051	0,134
X_9 Nivel de fidelidad	1,377	0,243	1,377	0,243	1,575	0,212	0,091	0,763	0,056	0,945
X_{10} Nivel de satisfacción	0,323	0,571	0,323	0,571	0,000	0,986	0,054	0,817	3,302	0,041

Nota: Los valores representan el valor y la significación estadística (Sig.) del test de Levene con la valoración de la dispersión de la varianza de cada variable métrica a través de los niveles de las variables no métricas/categóricas.

Los test de homocedasticidad de dos variables métricas, que encontramos en los métodos como el de regresión múltiple, se realizan mejor a través del análisis gráfico, particularmente un análisis de los residuos. Se remite al lector interesado al Capítulo 4 para una discusión completa del análisis de los residuos y los modelos de los residuos indicativos de heterocedasticidad.

Linealidad

El supuesto final a examinar es la linealidad de las relaciones. En el caso de variables individuales, se relacionan las pautas de asociación entre cada par de variables y la capacidad del coeficiente de correlación para representar adecuadamente la relación. Si hay un indicio de relaciones no lineales, entonces el investigador puede, tanto transformar una o ambas variables para conseguir la linealidad, como crear variables adicionales para representar los componentes no lineales. Nos podemos apoyar en la inspección visual de las relaciones para determinar si están presentes relaciones no lineales. El lector puede ir a la Figura 2.3, que presenta los gráficos de dispersión para todas las variables métricas en el conjunto de datos. El examen de los gráficos de dispersión no revela ninguna relación aparente. Por tanto, las transformaciones no se consideran necesarias. El supuesto de linealidad se comprobará también para el modelo multivariante entero y cunado se lleve a cabo el examen de los residuos en la regresión múltiple.

Resumen

Las series de test gráficos y estadísticos dirigidos hacia la evaluación de los supuestos subyacentes en las técnicas multivariantes descubre realmente poco en términos de violaciones de los supuestos. Donde se han indicado violaciones, éstas han sido relativamente pequeñas y no presentarían ningún problema serio para el análisis de datos posterior. Se anima al investigador a realizar siempre estos exámenes de los datos simples, aunque reveladores, para asegurar que puedan identificarse y resolverse los problemas potenciales antes de comenzar el análisis.

La incorporación de datos no métricos con variables ficticicas

Un factor crítico en la elección y aplicación de la técnica multivariante correcta es la medición de las propiedades de las variables dependientes e independientes. Algunas técnicas, tales como el análisis discriminante o el análisis multivariante de la varianza, requieren especificamente datos no métricos como variables dependientes o independientes. Sin embargo, en muchos casos, las variables métricas tienen que ser utilizadas como variables independientes, como ocurre en el análisis de regresión, en el análisis discriminante y en la correlación canónica. Además, las técnicas de interdependencia de análisis factorial y cluster normalmente requieren variables métricas. Con este fin, todos los debates han asumido la medición métrica de variables. Pero, ¿qué podemos hacer cuando las variables son no métricas, con dos o más categorías? ¿Se excluyen en muchas técnicas multivariantes las variables no métricas tales como género, situación marital u ocupación? La respuesta es negativa. A continuación, discutiremos como incorporar las variables no métricas a muchas de estas situaciones que requieren variables métricas.

El investigador tiene a su disposición un método para usar variables dicotómicas, conocidas como **variables ficticias**, que actúan como variables de sustitución. Una variable ficticia es una variable dictómica que representa una categoría de variable independiente no métrica. Cualquier variable no métrica con k categorías puede ser representada como variable ficticia $k - 1$. El siguiente ejemplo ayudará a clarificar este concepto.

En primer lugar, supongamos que queremos incluir el género, que tiene dos categorías, femenina y masculina. También hemos medido el nivel de renta doméstica con tres categorías (véase

TABLA 2.13. Representación de variables no métricas con variables ficticias

Variable no métrica con dos categorías (género)		*Variable no métrica con tres categorías (nivel de renta doméstica)*	
Género	*Variables ficticias*	*Nivel de renta doméstica*	*Variables ficticias*
Femenino	$X_1 = 1$, o bien $X_1 = 0$	si < 15.000$	$X_3 = 1$, o bien $X_3 = 0$
Masculino	$X_2 = 1$, o bien $X_2 = 0$	si ≥ 15.000$ & ≤25.000 $	$X_4 = 1$, o bien $X_4 = 0$
		si > 25.000$	$X_5 = 1$, o bien $X_5 = 0$

Tabla 2.13). Para representar el género no métrico de la variable, crearemos dos nuevas variables ficticias (X_1 y X_2), tal y como se muestra en la Tabla 2.13. X_1 representa a aquellos individuos de sexo femenino con un valor de 1, y dará a todos los varones un valor de 0. Del mismo modo, X_2 representa a todos los varones con el valor de 1 y da a todas las mujeres el valor de 0. Ambas variables, (X_1 y X_2) no son necesarias, no obstante, porque cuando $X_1 = 0$, el género tiene que ser femenino por definición. Por tanto, necesitamos incluir sólo una de las variables (X_1 o X_2) para probar el efecto del género.

De la misma forma, si medimos también la renta doméstica con tres niveles tal y como se muestra en la Tabla 2.13, nuestra primera definición serían tres variables ficticias (X_3, X_4 y X_5). Sin embargo, tal y como ocurre en el caso de género, no necesitamos el conjunto entero de variables ficticias, y en su lugar usamos variables ficticias $k - 1$, donde K es el número de categorías. Por tanto, usaremos dos de las variables ficticias para representar los efectos de la renta doméstica.

Existen tres maneras de representar los niveles de renta doméstica con dos variables ficticias, tal y como se muestra en la Tabla 2.14. Esta forma de codificación de variables ficticias se conoce como **codificación de indicadores**. Una consideración importante en esta forma de codificación de variables ficticias es recordar la categoría que se omite, que se conoce como **grupo de comparación**. Esta es la categoría que recibe los ceros para las variables ficticias. Por ejemplo, en el análisis de regresión, los coeficientes de regresión para las variables ficticias representan desviaciones del grupo de comparación en cuanto a la variable de criterio. Las desviaciones representan las diferencias entre las medias para cada grupo de encuestados compuesto por una variable ficticia y un grupo de comparación. Esta forma es más apropiada cuando existe un grupo de comparación lógica, como en el caso de un experimento. En un experimento con el grupo de control que actúa como el grupo de comparación, los coeficientes son las diferencias de media sobre la variable dependiente para cada grupo tratamiento del grupo de control. Cada vez que se usa la codificación de variables ficticias, tenemos que ser conscientes del grupo de comparación y recordar los impactos que tiene en nuestra interpretación de las variables restantes.

Un método alternativo de codificación de variables ficticias se denomina **codificación de efectos**. Es lo mismo que la codificación de indicadores excepto que el grupo de comparación (el grupo que recibe todos los ceros en la codificación de indicadores) recoge el valor de –1 en vez de 0 para las variables ficticias. Ahora los coeficientes representan las diferencias de cualquier grupo respecto a la media de todos los grupos en vez de a la del grupo omitido. Ambas formas de codificación de variables darán los mismos resultados; las únicas diferencias tendrán lugar en la interpretación de los coeficientes de las variables ficticias.

TABLA 2.14. Pautas de codificación de variables ficticias alternativas para variables no métricas de tres categorías

Nivel de renta doméstica	*Pauta 1*		*Pauta 2*		*Pauta 3*	
	X_3	X_4	X_3	X_4	X_3	X_4
Si < 15.000$	1	0	1	0	0	0
Si ≥ 15.000$ & ≤ 25.000$	0	1	0	0	1	0
Si > 25.000$	0	0	0	1	0	1

La utilización más frecuente de variables ficticias se produce en el análisis de regresión y discriminante, donde los coeficientes tienen una interpretación directa. Su empleo en otras técnicas multivariantes es más limitado, especialmente en las que dependen de pautas correlacionales, tales como el análisis factorial, porque la correlación de una variable binaria no queda muy bien representada por la correlación de coeficientes tradicional de Pearson. No obstante, se pueden hacer consideraciones especiales en estos casos, tal y como se indica en los capítulos correspondientes.

Resumen

Este capítulo ha proporcionado al investigador el instrumental necesario para examinar y explorar la naturaleza de los datos y las relaciones entre variables antes de la aplicación de cualquiera de las técnicas multivariantes. Aunque este análisis requieren una cantidad considerable de tiempo y esfuerzo, el investigador prudente invierte sabiamente los recursos necesarios en un profundo examen de los datos para asegurarse de que se aplican los métodos multivariantes en las situaciones apropiadas y para evaluar de forma más adecuada la interpretación de los resultados.

Cuestiones de revisión

1. Enumere las causas potenciales de existencia de casos atípicos. Asegúrese de incluir los atribuidos tanto al investigador como al encuestado.
2. Discuta por qué los casos atípicos pueden clasificarse como beneficiosos y problemáticos.
3. Distinga entre los procesos aleatorios de datos ausentes (MAR) y los completamente aleatorios (MCAR). Explique cómo influye cada tipo de proceso en el análisis de los datos ausentes.
4. Describa las condiciones bajo las cuales un analista eliminaría un caso con datos ausentes y las condiciones bajo las cuales un analista utilizaría un método de imputación.
5. Evalúe la siguiente afirmación: para utilizar la mayoría de los análisis multivariantes no es necesario que se cumplan todos los supuestos de normalidad, linealidad, homocedasticidad e independencia.
6. Discuta la siguiente afirmación: los análisis multivariantes puede realizarse sobre cualquier conjunto de datos, en la medida en que el tamaño de la muestra sea el adecuado.

Bibliografía

1. Anderson, Edgar (1969), «A Semigraphical Method for the Analysis of Complex Problems.» *Technometrics* 2 (August): 387–91.
2. Box, G. E. P., and D. R. Cox (1964), «An Analysis of Transformations.» *Journal of the Royal Statistical Society* B (26): 211–43.
3. Chernoff, Herman. «Graphical Representation as a Discipline,» in *Graphical Representation of Multivariate Data*, Peter C. C. Wang, ed. New York: Academic Press, pp. 1–11.
4. Cohen, Jacob, and Patricia Cohen (1983), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2d ed. Hillsdale, N.J.: Lawrence Erlbaum Associates.
5. Daniel, C., and F. S. Wood (1980), *Fitting Equations to Data*, 2d ed. New York: Wiley-Interscience.
6. Dempster, A. P., and D. B. Rubin (1983), «Overview,» in *Incomplete Data in Sample Surveys: Theory and Annotated Bibliography*, vol. 2. Madow, Olkin, and Rubin, eds. New York: Academic Press.
7. Feinberg, Stephen (1979), «Graphical Methods in Statistics.» *American Statistician* 33 (November): 165–78.
8. Johnson, R. A., and D. W. Wichern (1982), *Applied*

Multivariate Statistical Analysis. Upper Saddle River, N.J.: Prentice-Hall.
9. Little, Roderick J. A., and Donald B. Rubin (1987), *Statistical Analysis with Missing Data.* New York: Wiley.
10. Wang, Peter C. C., ed. (1978), *Graphical Representation of Multivariate Data.* New York: Academic Press.
11. Weisberg, S. (1985), *Applied Linear Regression.* New York: Wiley.
12. Wilkinson, L. (1982), «An Experimental Evaluation of Multivariate Graphical Point Representations.» In *Human Factors in Computer Systems: Proceedings*, New York: ACM Press, pp. 202–9.

CAPÍTULO 3

Análisis factorial

OBJETIVOS DE APRENDIZAJE

- Diferenciar las técnicas de análisis factorial de otras técnicas multivariantes.
- Determinar los usos principales de las técnicas de análisis factorial.
- Distinguir entre los usos exploratorios y confirmatorios de las técnicas de análisis factorial.
- Identificar las diferencias entre el análisis de componentes principales y los modelos de análisis factorial común.
- Saber cuándo usar el análisis de componentes principales y el análisis factorial común.
- Identificar la diferencia entre el análisis factorial R y Q.
- Explicar el concepto de rotación de factores.
- Saber cómo determinar el número de factores a extraer.
- Explicar cómo denominar a un factor.
- Entender las ventajas y desventajas asociadas con las escalas aditivas.
- Explicar el propósito de las puntuaciones de factores y cómo usarlas.
- Explicar cómo seleccionar las variables obtenidas para un análisis posterior.
- Determinar las restricciones de las técnicas de análisis factorial.

La técnica estadística multivariante del análisis factorial ha experimentado una utilización creciente durante la última década en todas las áreas de investigación de carácter empresarial. Conforme aumenta el número de variables necesarias en las técnicas multivariantes, existe una necesidad mayor de conocer en profundidad la estructura y las interrelaciones de las variables. Este capítulo describe el análisis factorial, una técnica especialmente adecuada para analizar las pautas de relaciones complejas y multidimensionales encontradas por los investigadores. Define y explica los aspectos fundamentales de las técnicas de análisis factorial en términos amplios y concep-

tuales. Se puede usar el análisis factorial para examinar las pautas subyacentes o las relaciones para un amplio número de variables y para determinar si la información puede ser condensada o resumida en una serie de factores o componentes más pequeños. Con el fin de hacer una clarificación más profunda de los conceptos metodológicos, también se incluyen directrices básicas para presentar e interpretar los resultados de estas técnicas.

¿Qué es el análisis factorial?

El análisis factorial es un nombre genérico que se da a una clase de métodos estadísticos multivariantes cuyo propósito principal es definir la estructura subyacente en una matriz de datos. Generalmente hablando, aborda el problema de cómo analizar la estructura de las interrelaciones (correlaciones) entre un gran número de variables (por ejemplo, las puntuaciones de prueba, artículos de prueba, respuestas de cuestionarios) con la definición de una serie de dimensiones subyacentes comunes, conocidas como **factores**. Con el análisis factorial, el investigador puede identificar primero las dimensiones separadas de la estructura y entonces determinar el grado en que se justifica cada variable por cada dimensión. Una vez que se determinan estas dimensiones y la explicación de cada variable, se pueden lograr los dos objetivos principales para el análisis factorial —el resumen y la reducción de datos. A la hora de resumir los datos, con el análisis factorial se obtienen unas dimensiones subyacentes que, cuando son interpretadas y comprendidas, describen los datos con un número de conceptos mucho más reducido que las variables individuales originales. Se puede obtener la reducción de datos con el cálculo de la puntuación para cada dimensión subyacente y sustituirlos por las variables originales.

Introducimos el análisis factorial como nuestra primera técnica multivariante porque puede jugar un papel único en la aplicación de otras técnicas multivariantes. Como ya hemos mencionado anteriormente, la ventaja principal de las técnicas multivariantes es su capacidad para acomodar las variables múltiples con el fin de comprender las relaciones complejas que no son posibles con los métodos univariantes y bivariantes. El incremento del número de las variables también aumenta la posibilidad de que las variables estén no correlacionadas y no sean representativas de unos conceptos distintos. En su lugar, los grupos de variables pueden estar interrelacionados en la medida en que son todos representativos de un concepto más general. Esto se puede deber al diseño, como es el intento de medir las muchas facetas de personalidad o imagen de establecimiento, o puede surgir simplemente de la adición de nuevas variables. En cualquier caso, el investigador tiene que saber cómo se relacionan las variables para interpretar mejor los resultados. Finalmente, si el número de variables es demasiado grande o existe la necesidad de dar una mejor representación a un número de conceptos más pequeño en vez de las facetas múltiples, el análisis factorial puede asistir en la selección de un subgrupo de variables representativo o incluso crear nuevas variables como sustitutas para las variables originales mientras mantengan su carácter original.

El análisis factorial es diferente de las técnicas de dependencia que se mencionan en la siguiente sección (es decir, la regresión múltiple, el análisis discriminante, el análisis multivariante de la varianza o la correlación canónica), en el que se consideran una o más variables explícitamente como las variables de criterio o dependientes y todas las demás son las variables de predicción o independientes. El análisis factorial es una técnica de interdependencia en la que se consideran todas las variables simultáneamente, cada una relacionada con todas las demás y empleando todavía el concepto del valor teórico, el compuesto lineal de las variables. En el análisis factorial, los valores teóricos (los factores) se forman para maximizar su explicación de la serie de variables entera, y no para predecir una(s) variable(s) dependiente(s). Si hiciéramos una analogía con las técnicas de dependencia, cada una de las variables (originales) observadas sería una variable dependiente, que es una función de una serie de factores (dimensiones) subyacentes y latentes que están compuestas por todas las otras variables. Por tanto, cada variable es predicha por todas las demás. Por el contrario, se puede considerar cada factor (valor teórico) como una variable dependiente que es

función del conjunto entero de las variables observadas. Cualquiera de estas analogías ilustra las diferencias de propósito entre las técnicas de dependencia (la predicción) y la interdependencia (identificación de estructura).

Las técnicas analíticas de factores pueden lograr sus propósitos desde una perspectiva exploratoria o confirmatoria. Existe una discusión continuada acerca del papel adecuado del análisis factorial. Muchos investigadores lo consideran meramente exploratorio, útil para la búsqueda de una estructura entre una serie de variables o como un método de reducción de datos. Desde esta perspectiva, las técnicas de análisis factorial «extraen lo que proporcionan los datos» y no tienen ninguna restricción a priori sobre la estimación de los componentes o el número de componentes a ser extraído. Para muchas aplicaciones, si no todas, resulta apropiada esta aplicación del análisis factorial. No obstante, en otras situaciones el investigador tiene unos pensamientos preconcebidos sobre la estructura real de los datos, que se basan en un apoyo teórico o investigaciones previas. Es posible que el investigador quiera probar las hipótesis que implican cuestiones tales como qué variables deberían ser agrupadas en un factor o el número exacto de factores. En estos casos, se requiere un análisis factorial que adopte un enfoque confirmatorio —es decir, valorar hasta qué punto los datos se ajustan a la estructura esperada. Los métodos tratados en este capítulo no proporcionan la estructura necesaria para la prueba de hipótesis formalizada. Abordamos explícitamente la perspectiva confirmatoria del análisis factorial en el Capítulo 11. En este capítulo, sin embargo, observamos las técnicas analíticas de factores principalmente desde un punto de vista exploratorio o no confirmatorio.

Un ejemplo hipotético del análisis factorial

Supongamos que mediante una investigación cualitativa un comercio minorista ha identificado ochenta características diferentes de comercios minoristas y su servicio, que los consumidores han mencionado que afectan su elección a la hora de frecuentar estos comercios. El comerciante quiere entender cómo deciden los consumidores, pero opina que no puede valorar las ochenta características individuales o desarrollar planes de acción para tantas variables, porque son demasiado específicos. En su lugar, al comercio le gustaría saber si los consumidores piensan en una dimensión determinante más general en vez de únicamente en aspectos específicos. Para identificar estas dimensiones, el comerciante podría encargar una encuesta solicitando valoraciones de consumidores sobre cada uno de estos aspectos específicos. Se emplearía entonces el análisis factorial para identificar las dimensiones determinantes subyacentes. Se considera que los aspectos específicos que se correlacionan en gran medida forman parte de una dimensión más amplia. Estas dimensiones se convierten en compuestos de las variables específicas, que a su vez permiten a las dimensiones ser interpretadas y descritas. En nuestro ejemplo, el análisis factorial podría identificar dimensiones como variedad de producto, calidad de producto, precios, personal de establecimiento, servicio y ambiente del establecimiento como las dimensiones determinantes utilizadas por los encuestados. Cada una de estas dimensiones contiene aspectos específicos que son una faceta de la dimensión determinante más amplia. A raíz de estos resultados, el comerciante puede usar estas dimensiones (factores) para definir áreas generales para la planificación y actuación.

Un ejemplo ilustrativo de una aplicación sencilla del análisis factorial se expone en la Figura 3.1, que representa la matriz de correlación para nueve elementos de imagen de un establecimiento. Se incluyen en esta serie las medidas de oferta del producto, personal del establecimiento, niveles de precios y el servicio y experiencia dentro del establecimiento. La pregunta que el investigador podrá resolver es: ¿están todos estos elementos separados en sus propiedades de valoración o están «agrupados» en áreas más generales de valoración? Por ejemplo, ¿se agrupan todos los elementos del producto juntos? ¿Dónde se ajusta el nivel de precios, o está separado? ¿Cómo se relacionan las características del funcionamiento dentro del establecimiento (por ejemplo, el personal del establecimiento, el servicio y el ambiente)? La inspección visual de la matriz de correlación original (Figura 3.1, parte 1) no revela fácilmente una pauta específica. Existen correlaciones dispersas

PARTE 1: MATRIZ DE CORRELACIÓN ORIGINAL

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
V_1 Nivel de precios	1,000								
V_2 Personal del establecimiento	0,427	1,000							
V_3 Política de retorno	0,302	0,771	1,000						
V_4 Disponibilidad del producto	0,470	0,497	0,427	1,000					
V_5 Calidad de producto	0,765	0,406	0,307	0,472	1,000				
V_6 Profundidad de surtido	0,281	0,445	0,423	0,713	0,325	1,000			
V_7 Anchura de surtido	0,354	0,490	0,471	0,719	0,378	0,724	1,000		
V_8 Servicio dentro del establecimiento	0,242	0,719	0,733	0,428	0,240	0,311	0,435	1,000	
V_9 Ambiente del establecimiento	0,372	0,737	0,774	0,479	0,326	0,429	0,466	0,710	1,000

PARTE 2: MATRIZ DE CORRELACIÓN DE LAS VARIABLES DESPUÉS DE AGRUPACIÓN SEGÚN EL ANÁLISIS FACTORIAL

	V_3	V_8	V_9	V_2	V_6	V_7	V_4	V_1	V_5
V_3 Política de retorno	1,000								
V_8 Servicio dentro del establecimiento	0,733	1,000							
V_9 Ambiente del establecimiento	0,774	0,710	1,000						
V_2 Personal del establecimiento	0,741	0,719	0,787	1,000					
V_6 Profundidad de surtido	0,423	0,311	0,429	0,445	1,000				
V_7 Anchura de surtido	0,471	0,435	0,468	0,490	0,724	1,000			
V_4 Disponibilidad del producto	0,427	0,428	0,479	0,497	0,713	0,719	1,000		
V_1 Nivel de precios	0,302	0,242	0,372	0,427	0,281	0,354	0,470	1,000	
V_5 Calidad de producto	0,307	0,240	0,326	0,406	0,325	0,378	0,472	0,765	1,000

Las áreas sombreadas representan las variables agrupadas por el análisis factorial.

FIGURA 3.1. Ejemplo ilustrativo del uso del análisis factorial para identificar la estructura dentro de un grupo de variables.

altas, pero las agrupaciones de variables no son evidentes. La aplicación del análisis factorial tiene como resultado la agrupación de variables tal y como se refleja en la parte 2 de la Figura 3.1. Aquí aparecen unas pautas interesantes. En primer lugar, se agrupan cuatro variables, todas relacionadas con la experiencia de compradores dentro del establecimiento. A continuación, tres variables que describen la variedad de producto y disponibilidad están agrupadas juntas. Finalmente, se agrupan la calidad de producto y los niveles de precios. Cada grupo representa una serie de variables altamente interrelacionadas que pueden reflejar una dimensión de valoración más general. En este caso, podríamos calificar las tres agrupaciones con la calificación de experiencia dentro del establecimiento, oferta de producto y valor. Todo esto proporcionaría a la dirección del establecimiento un número de conceptos más pequeño a tener en cuenta a la hora de formular la planificación comercial de estratégica o táctica, a la vez que proporciona un panorama más detallado de lo que constituye cada área general.

El proceso de decisión del análisis factorial

Centramos el debate del análisis factorial en el paradigma de construcción de modelos de seis pasos que se introdujo en el Capítulo 1. La Figura 3.2 muestra los primeros tres pasos de la aproximación estructurada para la construcción de modelos multivariantes, y la Figura 3.4 muestra en detalle los últimos tres pasos, y un paso adicional (el séptimo paso) más allá de la estimación, la interpretación y la validación de los modelos factoriales, que ayuda a la selección de las variables sustitutas, las puntuaciones de factores informáticos o la creación de las escalas aditivas para la utilización en otras técnicas multivariantes. A continuación se presenta un análisis de cada paso.

Primer paso: Objetivos del análisis factorial

El punto de comienzo en el análisis factorial, como con otras técnicas estadísticas, es el problema objeto de investigación. El propósito general de las técnicas analíticas de factores es encontrar una manera de condensar (resumir) la información contenida en una serie de variables originales en una serie más pequeña de dimensiones compuestas o valores teóricos (factores) nuevos con una mínima pérdida de información —es decir, buscar y definir las construcciones fundamentales o dimensiones que se supone sirven de base para las variables originales [17, 32]. Más específicamente, las técnicas del análisis factorial pueden satisfacer cualquiera de estos dos objetivos: (1) la identificación de estructura mediante el resumen de datos, o bien (2) la reducción de datos.

La identificación de estructura mediante el resumen de datos

El análisis factorial puede identificar la estructura de las relaciones entre las variables o los encuestados mediante la investigación de las correlaciones entre las variables o bien las correlaciones entre los encuestados. Por ejemplo, supongamos que tenemos datos sobre 100 encuestados basados en 10 características. Si el objetivo de la investigación fuera el resumen de las características, se aplicaría el análisis factorial a una matriz de correlación de las variables. Éste es el tipo de análisis factorial más común, y se denomina el análisis factorial R. **El análisis factorial R** analiza una serie de variables para identificar las dimensiones que son latentes (las que no son fácilmente observadas). También se puede aplicar el análisis factorial a la matriz de correlación de los encuestados individuales basada en sus características. Esto se denomina **el análisis factorial Q**, siendo un método para combinar o condensar grandes grupos de personas en grupos claramente diferentes dentro de una población mayor, no se usa la aproximación del análisis factorial Q con mucha frecuencia (dadas las dificultades informáticas). En su lugar, la mayoría de los investiga-

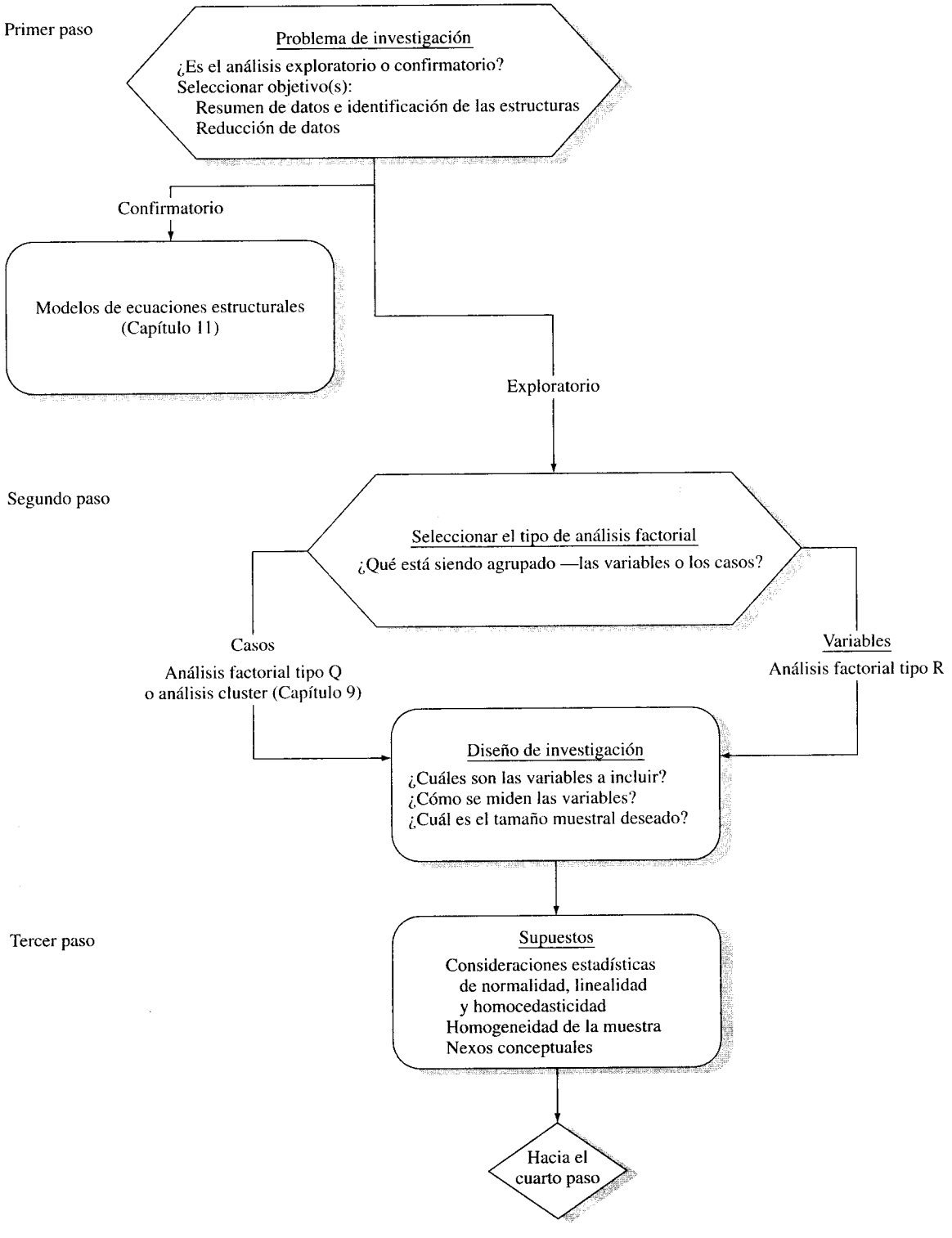


FIGURA 3.2. Pasos 1-3 en el diagrama de la decisión del análisis factorial.

dores utilizan algún tipo de **análisis cluster** (véase Capítulo 9) para agrupar los encuestados individuales. También véase Stewart [37] para otras combinaciones posibles de grupos y tipos de variables.

Reducción de datos

El análisis factorial también puede: (1) identificar las variables suplentes de una serie de variables más grande para su utilización en análisis multivariantes posteriores, o (2) crear una serie de variables completamente nueva, mucho más pequeña en número, para reemplazar parcial o completamente la serie original de variables para su inclusión en técnicas posteriores. En ambos casos, el propósito es retener la naturaleza y el carácter de las variables originales, pero reducir su número para simplificar el análisis multivariante posterior. Aunque las técnicas multivariantes se han desarrollado para utilizar múltiples variables, el investigador siempre busca la serie de variables más reducida para incluirla en el análisis. Tal y como se presentó en el Capítulo 1, tanto las cuestiones conceptuales como las empíricas respaldan la creación de medidas compuestas. El análisis factorial proporciona la base empírica para valorar la estructura de las variables y la potencial para crear estas medidas compuestas o seleccionar una subserie de variables suplentes para el análisis posterior.

El resumen de datos hace que la identificación de las dimensiones subyacentes o los factores sean fines de por sí; las estimaciones de los factores y las contribuciones de cada variable a los factores (denominadas cargas de los factores) constituyen todo lo que se necesita para el análisis. La reducción de datos depende también de las cargas de los factores; no obstante, las utiliza como la base para identificar las variables para su análisis subsiguiente con otras técnicas o bien para hacer estimaciones de los factores mismos (puntuaciones de factores o escalas aditivas), que a su vez reemplazan las variables originales en análisis subsiguientes. Más adelante, se trata el método de calcular e interpretar las cargas factoriales.

El uso del análisis factorial con otras técnicas multivariantes

El análisis factorial proporciona una visión directa de las interrelaciones entre las variables o los encuestados y un apoyo empírico para abordar las cuestiones conceptuales que tienen relación con la estructura subyacente de los datos. También juega un papel complementario importante con otras técnicas multivariantes mediante el resumen y la reducción de datos. Desde la perspectiva del resumen de datos, el análisis factorial proporciona al investigador una comprensión clara de cuáles de las variables podrían actuar juntas y cuántas de las variables realmente se puede esperar que tengan un impacto en el análisis. Por ejemplo, se esperaría que las variables altamente correlacionadas y miembros del mismo factor tuvieran perfiles similares de diferencia a través de los grupos en el análisis multivariante de la varianza o en el análisis discriminante. Los procedimientos que muestran el impacto de las variables correlacionadas son los basados en etapas (stepwise) de la regresión múltiple o el análisis discriminante. Estas técnicas introducen las variables, de forma secuencial, basadas en su capacidad adicional de predicción sobre las variables en el modelo. Conforme entra la variable de un factor, es menos probable que variables adicionales del mismo factor sean también incluidas, porque están altamente correlacionadas y potencialmente tienen menos capacidad de predicción adicional, que las variables que no estén en ese factor. Esto no significa que las otras variables del factor sean menos importantes o tengan menos impacto, sino que su efecto ya ha sido representado por la variable incluida en ese factor. El investigador comprenderá mejor el razonamiento que se halla tras la inclusión de variables en esta técnica con un conocimiento de la estructura de las variables.

La visión proporcionada por el resumen de datos puede ser incorporada directamente a otras técnicas multivariantes mediante cualquiera de las técnicas de reducción de datos. El análisis factorial proporciona la base para crear una nueva serie de variables que incorporan el carácter y naturaleza de las variables originales en una cantidad de nuevas variables más reducida, sea con la utilización de variables suplentes, sea con la puntuación de factores o las escalas aditivas. De esta manera, se pueden reducir los problemas que se asocian con las grandes cantidades de variables o intercorrelaciones altas entre las variables con la sustitución de las nuevas variables. El investigador puede beneficiarse de las relaciones y la visión detallada de la base conceptual y la interpretación de los resultados.

Selección de variables

La reducción y el resumen de datos pueden ser llevados a cabo tanto con una serie de variables preexistentes como con las variables creadas por la nueva investigación. Cuando se usa una nueva serie, el investigador debe realizar una aproximación conceptual para determinar qué variables conviene incluir en el análisis. El uso del análisis factorial para la reducción de datos es particularmente crítico cuando se requiere la comparabilidad a lo largo de un período de tiempo o en situaciones múltiples. Cuando se usa en una nueva investigación, el análisis factorial puede determinar también la estructura y/o crear nuevas puntuaciones compuestas a partir de las variables originales. Por ejemplo, uno de los primeros pasos en la construcción de la escala aditiva (véase Capítulo 1) es valorar la naturaleza de su dimensión y la conveniencia de las variables seleccionadas mediante el análisis factorial. Por tanto, aunque no es verdaderamente confirmatorio, el análisis factorial se puede utilizar para valorar la naturaleza de una dimensión propuesta.

Una vez que se especifica el propósito del análisis factorial, el investigador tiene que definir la serie de variables a examinar. Por lo que se refiere tanto al análisis factorial tipo R o tipo Q, el investigador especifica implícitamente las dimensiones potenciales que se pueden identificar mediante el carácter y la naturaleza de las variables sujetas al análisis factorial. Por ejemplo, en la valoración de las dimensiones de la imagen del establecimiento, el análisis factorial no podría identificar esta dimensión si no han sido incluidas preguntas sobre el personal del establecimiento. El investigador también tiene que recordar que el análisis factorial siempre producirá factores. Por tanto, el análisis factorial es siempre un candidato potencial para el fenómeno «basura dentro, basura fuera». Si el investigador incluye indiscriminadamente grandes cantidades de variables y espera que el análisis factorial «lo solucione», entonces la posibilidad de obtener malos resultados es alta. La calidad y el significado de los factores derivados reflejan un acercamiento conceptual a las variables incluidas en el análisis. El uso del análisis factorial como una técnica de resumen de datos no excluye la necesidad de una base conceptual para cualquiera de las variables analizadas. Incluso si se usa meramente para la reducción de datos, el análisis factorial es más eficiente cuando las dimensiones definidas conceptualmente pueden ser representadas por los factores obtenidos.

Segundo paso: El diseño de un análisis factorial

El diseño de un análisis factorial implica tres decisiones básicas: (1) el cálculo de los datos de entrada (una matriz de correlación) para alcanzar los objetivos específicos de la agrupación de variables o encuestados; (2) el diseño del estudio en términos de número de variables, las propiedades de medición de las variables y los tipos de las variables permisibles; y (3) el tamaño de muestra necesario, tanto en términos absolutos como para la función del número de variables en el análisis.

Las correlaciones entre las variables o los encuestados

La primera decisión en el diseño de un análisis factorial se concentra en la aproximación que se usa para calcular la matriz de correlación tanto para el análisis factorial de tipo R como para el de tipo Q. El investigador puede utilizar la matriz de datos de entrada a partir del cálculo de las correlaciones entre las variables, empleando, por tanto, un análisis factorial de tipo R. El investigador también puede elegir la matriz de correlación de las correlaciones entre los encuestados individuales. En este tipo de análisis factorial tipo Q, el resultado será una matriz factorial que identifica individuos similares. Por ejemplo, si los encuestados individuales se identifican por un número, la pauta de factores de resultado podría indicarnos que los individuos 1, 5, 6 y 7 son similares. Del mismo modo, los encuestados 2, 3, 4 y 8 quizá cargarían juntos sobre otro factor, y clasificaríamos estos individuos como similares. A partir de los resultados del análisis factorial Q, se pueden identificar grupos o clusters de individuos que muestran una pauta similar sobre las variables que se incluyen en el análisis.

	Variables		
Encuestados	V_1	V_2	V_3
A	7	7	8
B	8	6	6
C	2	2	3
D	3	1	1

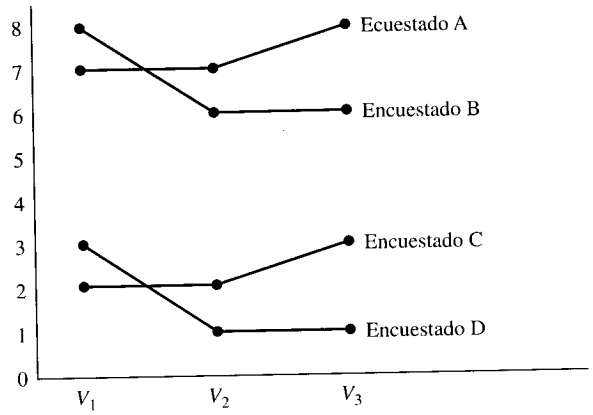


FIGURA 3.3. Comparaciones de perfiles de puntuaciones para el análisis factorial tipo Q y el análisis cluster.

A estas alturas, una pregunta lógica sería, ¿cómo se diferencia el análisis factorial tipo Q del análisis cluster, dado que ambas aproximaciones comparan la pauta de respuestas a través de una serie de variables y clasifican a los encuestados en grupos? La respuesta es que el análisis factorial tipo Q se basa en las intercorrelaciones entre los encuestados, mientras que el análisis cluster forma grupos que se basan en una medida de similitud basada en la distancia entre las puntuaciones de los encuestados sobre las variables analizadas. Para ilustrar esta diferencia, consideremos la Figura 3.3, que contiene las puntuaciones de cuatro encuestados sobre tres variables diferentes. Un análisis factorial tipo Q de estos cuatro encuestados daría dos grupos con estructuras de covarianza similares, agrupando a los encuestados A y C frente a B y D. Por contraste, la aproximación de cluster sería sensible a las distancias reales entre las puntuaciones de los encuestados y llevaría a la agrupación de las parejas más cercanas. Por tanto, con la aproximación del análisis cluster, los encuestados A y B estarían situados en un grupo y C y D en el otro grupo. Si el investigador decide emplear el análisis factorial tipo Q, debe ser consciente de estas diferencias. Con la disponibilidad de otras técnicas de agrupación y el uso general del análisis factorial para la reducción de datos y el resumen, el debate restante de este capítulo se centra en el análisis factorial tipo R, la agrupación de variables en vez de la agrupación de encuestados.

La selección de variables y cuestiones de medición

Ahora es necesario abordar dos preguntas: (1) ¿cómo se miden las variables? Y (2) ¿cuántas variables deberían ser incluidas? Se supone, por regla general, que las variables a incluir en el análisis factorial tienen escala métrica. En algunos casos, se pueden utilizar **variables ficticias** (codificadas 0-1), aunque se consideran como no métricas. Si todas las variables son ficticias, entonces las formas especializadas del análisis factorial, tales como el análisis factorial Boolean, son más apropiadas [5]. Además, el investigador debe intentar minimizar el número de variables que se incluyen; no obstante, debe mantener un número razonable de variables por factor. Si se está diseñando un estudio para valorar una estructura propuesta, el investigador deberá incluir varias variables (cinco o más) que puedan representar cada factor propuesto. El poder del análisis factorial se basa en encontrar pautas entre grupos de variables y resulta de poca utilidad en la identificación de factores compuestos de una única variable. Finalmente, cuando se diseña una investigación que utiliza análisis factorial, el investigador debería, si cabe, identificar varias variables claves (a veces denominadas indicadores claves o variables marcadoras) que reflejan con detalle los factores subyacentes hipotéticos, de forma que se facilite la validación de los factores derivados y la valoración sobre la significación práctica de los resultados.

Tamaño muestral

En cuanto a la pregunta de tamaño muestral, generalmente el investigador no usará el análisis factorial para una muestra inferior a 50 observaciones, y preferiblemente el tamaño muestral debería ser 100 o más grande. Como regla general, el mínimo es tener por lo menos un número de observaciones cinco veces mayor que el número de variables a ser analizadas, siendo el tamaño aceptable un ratio de diez a uno. Algunos investigadores proponen incluso un mínimo de 20 casos por cada variable. Hay que recordar, sin embargo, que con 30 variables, por ejemplo, hay 435 correlaciones en el análisis factorial. Con un nivel de significación de 0,05, es posible que incluso 20 de estas correlaciones sean consideradas significativas y aparecerían en el análisis factorial simplemente por casualidad. El investigador siempre debería procurar obtener el ratio más alto de casos por variable para minimizar las posibilidades de «sobreajustar» los datos (es decir, derivar los factores que son específicos a la muestra con poca capacidad de generalización). De todas formas, se emplea una serie de variables menor al estar guiado por consideraciones conceptuales y prácticas. Aun así, se tienen tamaños muestrales más pequeños y/o ratios más bajos de casos y variables, debiéndose interpretar los resultados con cautela. La cuestión del tamaño muestral será abordada también en una sección posterior sobre la interpretación de las cargas de los factores.

Tercer paso: Supuestos en el análisis factorial

Los supuestos básicos subyacentes del análisis factorial son más de tipo conceptual que estadístico. Desde un punto de vista estadístico, se pueden obviar los supuestos de normalidad, homocedasticidad y linealidad siendo conscientes de que su incumplimiento produce una disminución en las correlaciones observadas. En realidad, sólo es necesaria la normalidad cuando se aplica una prueba estadística a la significación de los factores; sin embargo, raramente se utilizan estas pruebas. De hecho, es deseable que haya cierto grado de **multicolinealidad,** dado que el objetivo es identificar series de variables interrelacionadas.

Adicionalmente a las bases estadísticas para las correlaciones de la matriz de los datos, el investigador tiene que asegurarse también de que la matriz tiene suficientes correlaciones para justificar la aplicación del análisis factorial. Si la inspección visual revela que no hay un número sustancial de correlaciones mayores que 0,30, entonces el análisis factorial es probablemente inapropiado. Las correlaciones entre las variables también pueden ser analizadas con el cálculo de las correlaciones parciales entre las variables; esto es, las correlaciones entre variables cuando se tienen en cuenta los efectos de las otras variables. Si los factores «verdaderos» existen en los datos, la correlación parcial será pequeña, porque se puede explicar la variable mediante los factores (valores teóricos con cargas para cada variable). Si las correlaciones parciales son altas, entonces no existen factores subyacentes «verdaderos», y el análisis factorial es inapropiado. SPSS [36] y SAS [33] proporcionan la **matriz de correlación anti-imagen**, que es simplemente el valor negativo de la correlación parcial, mientras que el BMDP 5 proporciona directamente las correlaciones parciales. En cada caso, las correlaciones parciales o anti-imagen mayores son indicativas de una matriz de datos que no es quizá adecuada para el análisis factorial.

Otra manera de determinar la conveniencia del análisis factorial es examinar la matriz de correlación entera. El **contraste de esfericidad de Bartlett,** una prueba estadística para la presencia de correlaciones entre las variables, es una de estas medidas. Proporciona la probabilidad estadística de que la matriz de correlación de las variables sea una matriz identidad. El investigador debe tener en cuenta, sin embargo, que el incremento del tamaño muestral da lugar a que la prueba de contraste de Bartlett sea más sensible a la detección de correlaciones entre las variables. Otra medida para cuantificar el grado de intercorrelaciones entre las variables y la conveniencia del análisis factorial es la **medida de suficiencia de muestreo (MSA)**. Este índice se extiende de 0 a 1, llegando a 1 cuando cada variable es perfectamente predicha sin error por las otras variables. La medida puede ser interpretada con las siguientes directrices: 0,80 o superior, sobresaliente; 0,70 o superior, regular; 0,60 o superior, mediocre; 0,50 o superior, despreciable; y por debajo de 0,50,

inaceptable [21, 22]. El MSA aumenta conforme: (1) aumenta el tamaño muestral, (2) aumentan las correlaciones medias, (3) aumenta el número de variables o (4) desciende el número de factores [22]. Las mismas directrices de MSA pueden extenderse también a las variables individuales. El investigador debería examinar primero los valores MSA para cada variable y excluir aquellos que caen en la gama de inaceptables. Una vez que las variables individuales logran un nivel aceptable, se puede valorar el MSA general y se puede tomar una decisión sobre la continuación del análisis factorial.

Los supuestos conceptuales que subyacen en el análisis factorial se relacionan con la serie de variables seleccionadas y la muestra elegida. Un supuesto básico del análisis factorial es que existe una estructura subyacente en la serie de variables seleccionadas. Es responsabilidad del investigador asegurarse de que las pautas observadas sean válidas y conceptualmente apropiadas para utilizar el análisis factorial porque la técnica no tiene medios de determinar la conveniencia, aparte de las correlaciones entre las variables. Por ejemplo, la mezcla de variables dependientes e independientes en un solo análisis factorial y posteriormente el uso de los factores derivados para apoyar las relaciones de dependencia es inapropiado. El investigador debería asegurarse también de que la muestra es homogénea con respecto a la estructura de factor subyacente. Por ejemplo, la aplicación del análisis factorial resultaría inapropiada para una muestra de hombres y mujeres que tienen distintas opiniones sobre una serie de aspectos según el sexo. Cuando se combinan las dos submuestras (hombres y mujeres), las correlaciones resultantes y la estructura de factores serán una representación pobre de la estructura única de cada grupo. Por tanto, cuando se esperan grupos diferentes en la muestra, se deben practicar análisis factoriales separados y los resultados deben ser comparados para identificar las diferencias que no se reflejan en los resultados de la muestra combinada.

Cuarto paso: La estimación de los factores y la valoración del ajuste general

Una vez que se especifican las variables y se prepara la matriz de correlación, el investigador está preparado para aplicar el análisis factorial que identifique la estructura subyacente de las relaciones (véase Figura 3.4). Para realizar esta operación, es necesario tomar decisiones con relación a (1) el método de extracción de los factores (análisis factorial común frente al análisis de componentes principales) y (2) el número de factores seleccionados para representar la estructura subyacente en los datos. La selección del método de extracción depende del objetivo del investigador. Se utiliza el análisis de componentes principales cuando el objetivo es resumir la mayoría de la información original (varianza) en una cantidad mínima de factores con propósitos de predicción. Por el contrario, se utiliza el análisis factorial común para identificar los factores subyacentes o las dimensiones que reflejan qué es lo que las variables comparten en común. Para cualquiera de estos métodos, el investigador tiene que determinar también el número de factores que representan la serie de variables originales. Tanto las cuestiones conceptuales como empíricas afectan a esta decisión.

El análisis factorial común frente al análisis de componentes principales

El investigador puede utilizar dos modelos básicos para obtener soluciones factoriales. Éstos se conocen como **análisis factorial común** y **análisis de componentes principales**. Con el fin de seleccionar el modelo apropiado, en primer lugar el investigador tiene que comprender las diferencias entre los tipos de varianza. Para los propósitos del análisis factorial, existen tres tipos de varianza total: (1) **común**, (2) **específica** (también conocida como única), y (3) **error**. Estos tipos de varianza y su relación con el proceso de selección de modelo factorial se ilustran en la Figura 3.5. Se define la varianza común como aquella varianza en una variable que se comparte con

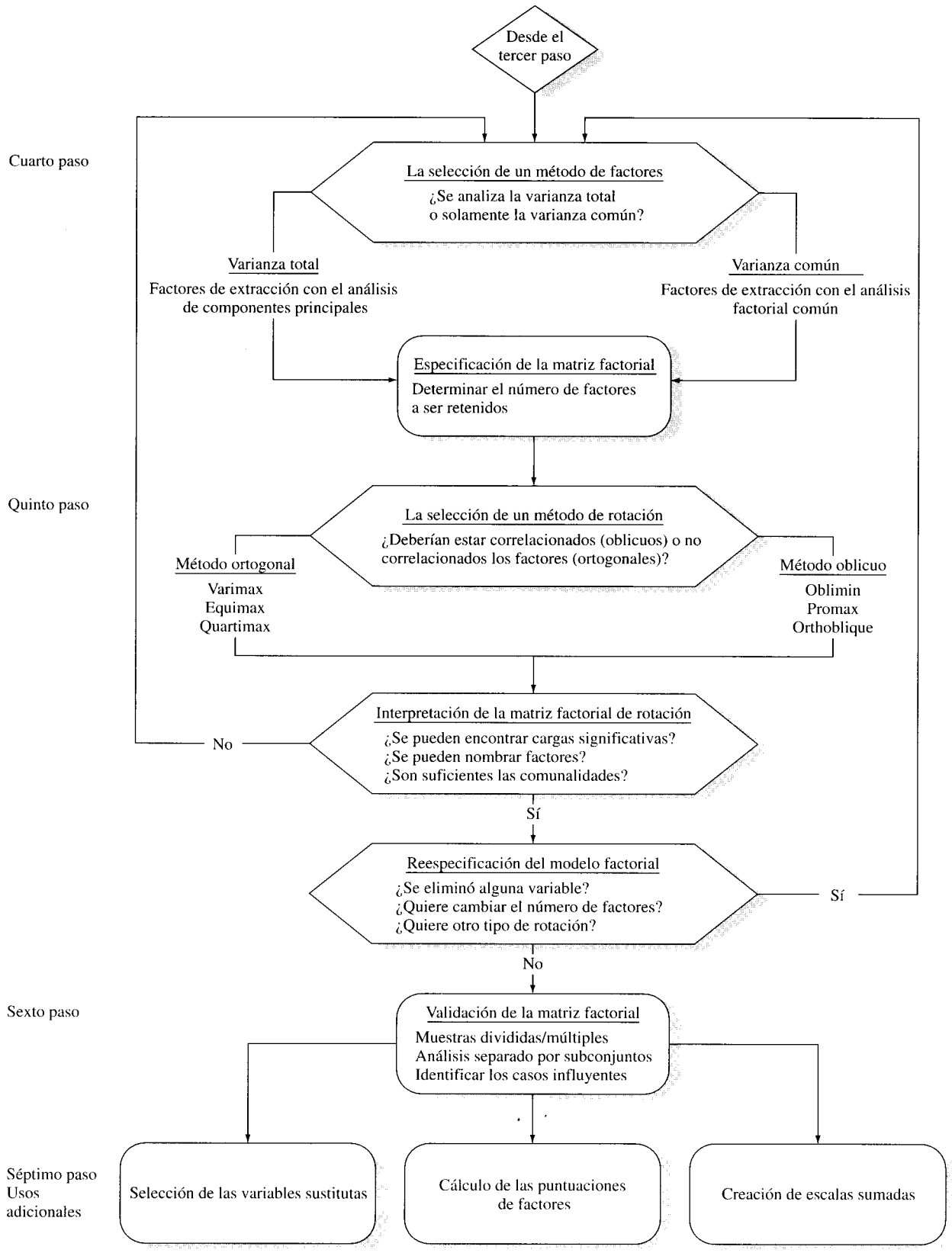


FIGURA 3.4. Pasos 4-7 en el diagrama de decisión del análisis factorial.

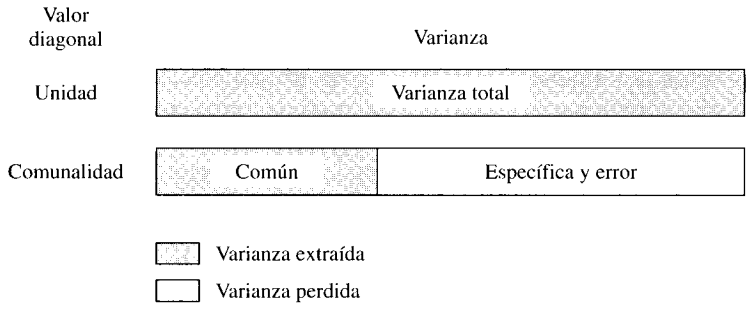


FIGURA 3.5. Tipos de varianza llevados a la matriz factorial.

todas las otras variables en el análisis. La varianza específica es aquella varianza asociada solamente con una variable específica. La varianza de error es aquella varianza que se debe a la poca fiabilidad en el proceso de recolección de datos, al error de medición o un componente aleatorio en el fenómeno medido. El análisis de componentes principales considera la varianza total y estima los factores que contienen proporciones bajas de la varianza única y, en algunos casos, la varianza de error. No obstante, los primeros factores no contienen la suficiente varianza única o de error como para distorsionar la estructura de factores en su conjunto. Específicamente, con el análisis de componentes principales, se insertan las unidades en la diagonal de la matriz de correlación, para que se traiga la varianza completa en la matriz de factores, tal y como se indica en la Figura 3.5. Por el contrario, en el análisis factorial común se incorporan las varianzas compartidas en la diagonal. **Las comunalidades** son estimaciones de la varianza compartida o común entre las variables. Los factores que resultan del análisis factorial común se basan solamente en la varianza común.

La selección de un modelo u otro se basa en dos criterios: (1) los objetivos del análisis factorial y (2) el grado de conocimiento anterior acerca de la varianza en las variables. El análisis de componentes principales es apropiado cuando el interés primordial se centra en la predicción o el mínimo número de factores necesarios para justificar la porción máxima de la varianza representada en la serie de variables original, y cuando el conocimiento previo sugiere que la varianza específica y de error representan una proporción relativamente pequeña de la varianza total. Por el contrario, cuando el objetivo principal es identificar las dimensiones latentes o las construcciones representadas en las variables originales y el investigador tiene poco conocimiento acerca de la varianza específica y de error y por tanto quiere eliminar esta varianza, lo más apropiado es utilizar el modelo factorial común. Se considera que el análisis factorial común, con unos supuestos más restrictivos y la utilización exclusiva de las dimensiones latentes (varianza compartida), se basa más en la teoría. Aunque teóricamente válido, no obstante, el análisis factorial común tiene varios problemas. En primer lugar, el análisis factorial común adolece de **indeterminación de factores**. Esto significa que para cualquier encuestado individual, se pueden calcular varias puntuaciones de factores diferentes a partir de los resultados del modelo factorial [25]. No existe una sola solución única, tal y como ocurre con el análisis de componentes principales; no obstante, y en la mayor parte de los casos, las diferencias no son sustanciales. La segunda cuestión se refiere al cálculo de las varianzas compartidas estimadas que se usan para representar la varianza compartida. Cuando se utilizan muestras de gran tamaño o un gran número de variables, los cálculos pueden requerir tiempo y recursos sustanciales del computador. Además, las varianzas compartidas no siempre se pueden estimar o pueden no ser válidas (es decir, valores mayores que 1 o menores que 0), lo que requiere la supresión de la variable del análisis (véase el ejemplo empírico del análisis factorial común abordado más adelante en este capítulo).

Las complicaciones del análisis factorial común han contribuido al uso generalizado del análisis de componentes principales. Aunque todavía los expertos siguen discutiendo acerca de cuál de los modelos factoriales es el más apropiado [6, 18, 24, 35], la investigación empírica ha mostrado resultados similares en muchos casos [38]. En la mayoría de las aplicaciones, tanto el análisis de componentes principales como los análisis factoriales comunes llegan a resultados

esencialmente idénticos si el número de variables excede de 30 [17], o las varianzas compartidas exceden de 0,60 para la mayoría de las variables. Si el investigador está preocupado por los supuestos del análisis de componentes principales, entonces debe aplicar también el análisis factorial común para valorar su estructura de representación.

Cuando se llega a una decisión acerca del modelo factorial, el investigador está preparado para extraer los factores sin rotación iniciales. Con el examen de la matriz de factores sin rotación, el investigador puede explorar las posibilidades de reducción de datos para una serie de variables y obtener una estimación preliminar de los factores a extraer. Sin embargo, se debe esperar para determinar el número de factores final hasta que se haga una rotación de los resultados y se interpreten los factores.

Criterios para el cálculo del número de factores a ser extraídos

¿Cómo decidimos el número de factores que se deben extraer? Cuando una gran serie de variables se somete a la extracción de factores, en primer lugar el método extrae las combinaciones de las variables que explican la cantidad mayor de la varianza y después continúa con combinaciones que justifican cantidades de varianza cada vez menores. Para decidir cuántos factores se deben extraer, el investigador empieza generalmente con algún criterio predeterminado, tal como el porcentaje de varianza o el criterio de raíz latente, para llegar a un número de factores específico (se abordan estas dos técnicas con más detalle más adelante). Después de estimar la solución inicial, se calculan varias soluciones de prueba adicionales —normalmente un factor menos que el número inicial y dos o tres factores más que los que se estimaron inicialmente. Posteriormente, en función de la información que se obtiene de estos análisis previos, se examinan las matrices de factores y se escoge el número de factores que represente mejor los datos. Por analogía, la elección del número de factores es como enfocar un microscopio. Un ajuste demasiado alto o demasiado bajo hará más oscura una estructura que es obvia cuando el ajuste es acertado. Por tanto, al examinar un número de estructuras factoriales diferentes que se derivan de varias soluciones de pruebas, el investigador puede comparar y contrastar para llegar a la mejor representación de los datos. Se puede decir que todavía no se ha desarrollado una base cuantitativa exacta para decidir el número de factores a extraer. No obstante, en general se utilizan los siguientes criterios para la extracción del número de factores.

Criterio de raíz latente La técnica que se utiliza más frecuentemente es el criterio de raíz latente. Esta técnica es sencilla de aplicar. La racionalidad que se usa para el criterio de raíz latente es que cualquier factor individual debería justificar la varianza de por lo menos una única variable. Cada variable contribuye con un valor de 1 para el autovalor total. Por tanto, sólo se consideran los factores que tienen **raíces latentes** o **autovalores** mayores que 1; explican al menos una variable, se considera que todos los factores con raíces latentes menores que 1 (explican menos de una variable) no son significativas y por tanto, se desestiman a la hora de incorporarlos a la interpretación. El uso del autovalor para establecer un corte es más fiable cuando el número de variables está entre 20 y 50. Si el número de variables es menor que 20, hay una tendencia a que este método extraiga un número de factores conservador (demasiado poco); por el contrario, si hay más de 50 variables, lo más común es que se extraigan demasiados factores.

Criterio a priori El criterio a priori es un criterio simple y a la vez razonable bajo ciertas circunstancias. Con su aplicación, el investigador ya sabe cuántos factores hay que extraer antes de iniciar el análisis factorial. El investigador simplemente instruye al computador para parar el análisis cuando se haya extraído el número de factores deseado. Esta aproximación resulta de utilidad cuando se prueba una teoría o una hipótesis acerca del número de factores para ser extraído. También puede ser justificado con el objetivo de replicar el trabajo de otro investigador y extraer el mismo número de factores que se encontraron previamente.

Criterio de porcentaje de la varianza El criterio de porcentaje de varianza es una aproximación que se basa en obtener un porcentaje acumulado especificado de la varianza total extraída. El propósito es asegurar una significación práctica de los factores derivados, asegurando que explican por lo menos una cantidad especificada de la varianza. No se ha adoptado un umbral absoluto para todas las aplicaciones. Sin embargo, en las ciencias naturales, el procedimiento de factores normalmente no debería ser detenido hasta que los factores extraídos cuenten por lo menos con un 95 por ciento de la varianza o hasta que el factor justifique solamente una porción pequeña (menos del 5 por ciento). Por contraste, en las ciencias sociales, donde la información muchas veces es menos precisa, es normal considerar una solución que represente un 60 por ciento de la varianza total (y en algunos casos incluso menos) como satisfactoria.

Una variante de este criterio implica la selección de suficientes factores para obtener una comunalidad para cada una de las variables. Si las razones teóricas o prácticas requieren una cierta varianza compartida para cada variable, entonces la investigación incluirá tantos factores como sean necesarios para representar de forma adecuada cada una de las variables originales.

Criterio de contraste de caída Recordemos que con el modelo de análisis de componentes principales, los factores posteriores que han sido extraídos contienen tanto la varianza común como la varianza única. Aunque todos los factores contienen por lo menos alguna varianza única, la proporción de la varianza única es sustancialmente más alta en los factores posteriores que en los primeros. El contraste de caída se utiliza para identificar el número óptimo de factores que pueden ser extraídos antes de que la cantidad de la varianza única empiece a dominar la estructura de la varianza común [9]. Se estima el contraste de caída con el trazo de raíces latentes en función del número de factores en su orden de extracción, y se utiliza la forma de la curva consiguiente para evaluar el punto de corte. La Figura 3.6 representa los primeros 18 factores extraídos de una investigación realizada por todos los autores de este libro. Si empezamos con el primer factor, el trazo tiene inicialmente una inclinación descendente y a continuación se convierte paulatinamente en una línea más o menos horizontal. Se considera que el punto en que la curva empieza a rectificarse indica el máximo número de factores a extraer. En el caso que nos ocupa, se incluirán los primeros 10 factores. Por encima de 10, se incluiría una proporción de la varianza única demasiado grande, por lo que estos factores no son aceptables. Es importante señalar que con el uso del criterio de raíz latente, solamente se tienen en cuenta ocho factores. Por el contrario, la utilización del criterio de contraste de caída nos proporciona dos factores más. Por regla general, el contraste de caída normalmente tiene como resultado que se incluyan uno y a veces dos o más factores *adicionales* que cuando se utiliza el criterio de raíz latente [9].

Heterogeneidad de la muestra La existencia de varianza compartida entre las variables es el núcleo tanto de los modelos de factores comunes como de los de componentes. Un supuesto subyacente es que la varianza compartida se extiende a lo largo de toda la muestra. Si la muestra es heterogénea al menos con respecto a un subconjunto de variables, los primeros factores representarán aquellas variables que son más homogéneas a lo largo de toda la muestra. Las variables con mayor capacidad de discriminar entre subconjuntos muestrales cargarán sobre los últimos factores, en muchas ocasiones aquellos no seleccionados de acuerdo a los criterios comentados más arriba [16]. Cuando el objetivo sea identificar factores que discriminen entre subconjuntos muestrales, el investigador deberá extraer factores adicionales entre aquellos señalados por los métodos anteriormente expuestos y examinar la capacidad de los factores adicionales para discriminar entre grupos. Si resultan ser peores al discriminar, la solución puede estar en proceder de nuevo y eliminar estos últimos factores.

Resumen de los criterios de selección de factores En la práctica, rara vez los investigadores utilizan un único criterio al determinar cuántos factores extraer. En su lugar, inicialmente emplean un criterio como el de la raíz latente como orientación en un primer intento

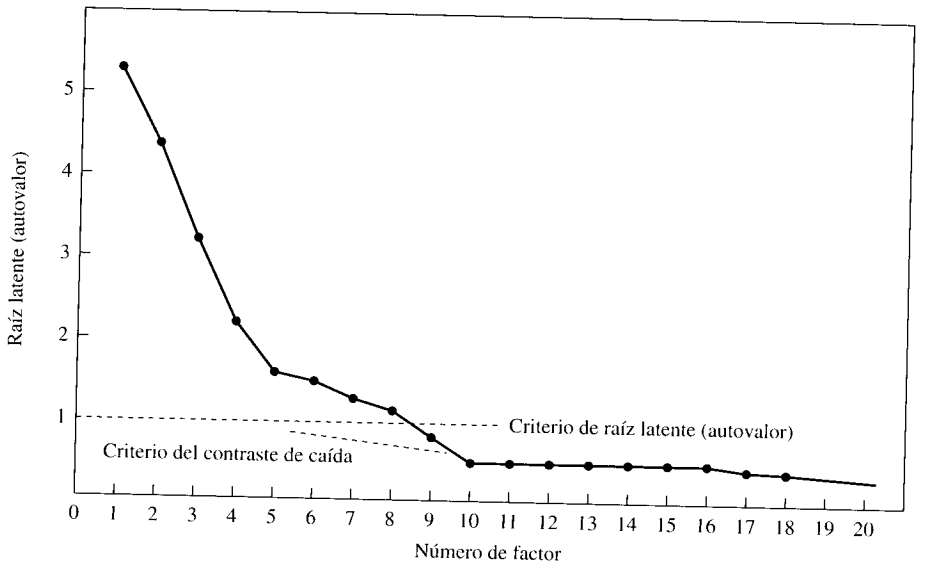


FIGURA 3.6. Gráfico de autovalor para el criterio de contraste de caída.

de interpretación. Después de haber interpretado los factores, como se expone en la siguiente sección, se valora su carácter práctico. También se interpretan los factores identificados mediante otros criterios. Elegir el número de factores está interrelacionado con la valoración de la estructura, lo que se revela en la etapa de interpretación. De esta forma, se examinan varias soluciones factoriales con diferentes números de factores antes de que la estructura esté bien definida.

Un aviso previo a la selección del conjunto definitivo de factores: puede resultar inconveniente seleccionar tanto muchos como pocos factores para representar los datos. Si se emplean pocos factores, no se revela la estructura correcta y pueden omitirse importantes dimensiones. Si se mantiene demasiados factores, las interpretaciones resultan más difíciles cuando se rotan los resultados (como se expone en la siguiente sección). Tal y como ocurre con otros aspectos de los modelos multivariantes, es importante la parsimonia. Una excepción a destacar es cuando el análisis de factores se emplea en exclusiva para la reducción de datos y se especifica la extracción de un nivel conjunto de varianza. El investigador debería siempre esforzarse en conseguir el conjunto de factores lo más representativo y parsimonioso posible.

Quinto paso: Interpretación de los factores

Para interpretar los factores y seleccionar la solución factorial definitiva se deben seguir tres pasos. En primer lugar, se calcula la **matriz inicial de factores** no rotados para que nos dé una indicación preliminar acerca del número de factores a extraer. La matriz de factores contiene las cargas factoriales (véase el siguiente párrafo) para cada variable sobre cada factor. Al calcular la matriz de factores no rotada, el investigador simplemente está interesado en la mejor combinación lineal de variables, es decir, en encontrar aquella combinación particular de las variables originales que cuenta con el mayor porcentaje de varianza de los datos. En consecuencia, el primer factor puede contemplarse como el mejor resumen de las relaciones lineales que los datos manifiestan. El segundo factor se define como la segunda mejor combinación lineal de las variables, sujeta a la restricción de que sea ortogonal al primer factor. Para ser **ortogonal** al primer factor, el segundo factor debe derivarse de la varianza restante tras la extracción del primer factor. Así, el segundo factor puede definirse como la combinación lineal de las variables que da cuenta del mayor porcentaje de varianza residual una vez se ha eliminado de los datos el efecto del primer factor. Los factores subsiguientes se definen de forma análoga hasta haber agotado la varianza de los datos.

Las soluciones factoriales no rotadas alcanzan el objetivo de reducción de los datos, pero el investigador debe preguntarse si la solución factorial no rotada (que se adecua a los requisitos matemáticos deseables) facilita una información que ofrezca la interpretación más adecuada de las variables examinadas. La mayor parte de las veces no resulta ser así. La carga factorial es el medio para interpretar la función que cada variable desempeña al definir cada factor. Las **cargas factoriales** son las correlaciones entre cada variable y el factor. Las cargas indican el grado de correspondencia entre cada variable y el factor, haciendo a una variable con mayor carga representativa del factor. La solución factorial no rotada puede no suministrar un patrón significativo de cargas de las variables. Si se espera que los factores no rotados sean significativos, el usuario puede especificar que la rotación no se lleve a cabo. Generalmente, la rotación es deseable porque simplifica la estructura de los factores, y habitualmente es difícil determinar si los factores no rotados serán significativos. Por tanto, el segundo paso hace uso de un método de rotación para lograr soluciones factoriales más simples y teóricamente más significativas. En muchos casos la rotación de los factores mejora la interpretación disminuyendo alguna de las ambigüedades que a menudo acompañan a las soluciones factoriales inicialmente no rotadas.

En una tercera etapa, el investigador valora la necesidad de especificar de nuevo el modelo de factores debido a (1) la eliminación de variables en el análisis, (2) el deseo de emplear un método de rotación diferente para la interpretación, (3) la necesidad de extraer un número diferente de factores, o (4) el deseo de cambiar de un método de extracción a otro. La especificación nueva del modelo factorial viene acompañada de la vuelta a la etapa de extracción, rotación de factores y de nuevo su interpretación.

Rotación de factores

Una herramienta importante al interpretar los factores es la **rotación de factores**. El término rotación significa exactamente lo que indica. Concretamente, se giran en el origen los ejes de referencia de los factores hasta alcanzar una determinada posición. Como se indicó previamente, las soluciones factoriales no rotadas extraen factores según su orden de importancia. El primer factor tiende a ser un factor general por el que casi toda variable se ve afectada significativamente dando cuenta del mayor porcentaje de varianza. El segundo y siguientes factores se basan en la varianza residual. Cada uno explica porcentajes de varianza cada vez menores. El efecto último de rotar la matriz de factores es redistribuir la varianza de los primeros factores a los últimos para lograr un patrón de factores más simple y teóricamente más significativo.

El caso más simple de rotación es la **rotación ortogonal**, en la que los ejes se mantienen formando un ángulo de 90 grados. También es posible rotar los ejes y no mantener el ángulo de 90 grados entre los ejes de referencia. Cuando no se limita a ser ortogonal, la rotación se denomina **oblicua**. Las rotaciones de factores ortogonal y oblicua están ilustradas en las Figuras 3.7 y 3.8 respectivamente.

Una ilustración de la rotación de factores La Figura 3.7, en el que se han representado cinco variables en un diagrama de factores bidimensional, ilustra la rotación de factores. El eje vertical representa el factor no rotado II, y el horizontal el factor no rotado I. El 0 indica el origen de coordenadas yendo éstas de –1,0 a 1,0. El número sobre los ejes representa las cargas factoriales. Las cinco variables están denominadas como V_1, V_2, V_3, V_4 y V_5. La carga factorial para la variable 2 (V_2) sobre el factor no rotado II está indicada horizontalmente mediante una línea de rayas del punto de los datos al eje vertical del factor II. Análogamente se dibuja una línea vertical de la variable 2 al eje horizontal del factor no rotado I para determinar la carga de la variable 2 sobre el factor 1. Un procedimiento similar para las variables restantes determina las cargas factoriales para las soluciones no rotadas y rotadas, como se muestra en la Tabla 3.1 a efectos comparativos. Sobre el primer factor no rotado, todas las variables cargan bastante alto. Sobre el segundo factor no rotado, las variables 1 y 2 cargan muy alto en el lado positivo.

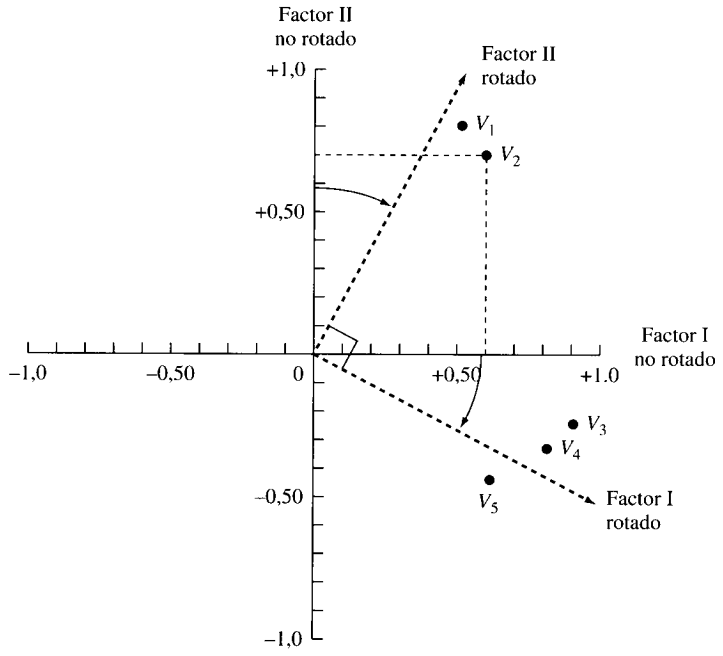


FIGURA 3.7. Rotación factorial ortogonal.

La variable 5 tiene una carga moderadamente alta en el lado negativo, y las variables 3 y 4 tienen cargas considerablemente inferiores en el lado negativo.

De la inspección visual de la Figura 3.7, resulta obvio que hay dos grupos de variables. Las variables 1 y 2 van juntas, así como las 3, 4 y 5. Sin embargo, este patrón de variables no es tan obvio a partir de las cargas de factores no rotados. Rotando los ejes originales en el sentido de las agujas del reloj, como se indica en la Figura 3.7, obtenemos un patrón de carga factorial completamente distinto. Nótese que al rotar los factores, los ejes mantienen el ángulo de 90 grados. Este

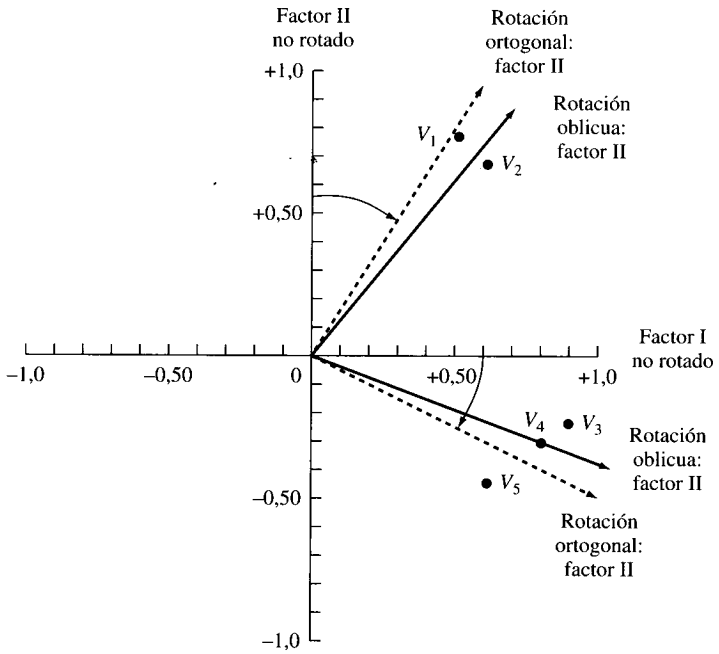


FIGURA 3.8. Rotación factorial oblicua.

TABLA 3.1. Comparación entre cargas factoriales rotadas y no rotadas

	Cargas factoriales no rotadas		*Cargas factoriales rotadas*	
Variables	*I*	*II*	*I*	*II*
V_1	0,50	0,80	0,03	0,94
V_2	0,60	0,70	0,16	0,90
V_3	0,90	–0,25	0,95	0,24
V_4	0,80	–0,30	0,84	0,15
V_5	0,60	–0,50	0,76	–0,13

procedimiento implica que los factores son matemáticamente independientes y que la rotación ha sido ortogonal. Después de rotar el eje de factores, las variables 3 y 4 cargan muy poco sobre el factor I, y las 1 y 2 cargan mucho sobre el factor II. Así, el patrón o agrupamiento de estas variables en dos grupos resulta más obvio que antes de la rotación, incluso la posición relativa o configuración de las variables permanece inalterada.

Los mismos principios generales de las rotaciones ortogonales atañen a las oblicuas. El método de la rotación oblicua es más flexible porque los ejes de factores no necesitan ser ortogonales. También es más realista porque las dimensiones subyacentes teóricamente más importantes, se suponen relacionadas entre sí. En la Figura 3.8 se comparan los dos métodos de rotación. Nótese que la rotación de factores oblicua representa el agrupamiento de variables con más precisión. Esa precisión se deriva del hecho de que cada eje de factores rotado está ahora más cerca del grupo respectivo de variables. Además, la solución oblicua provee de información sobre la medida en que los factores realmente están correlacionados uno con otro.

La mayor parte de los investigadores están de acuerdo en que soluciones factoriales no rotadas, aunque más directas, no resultan suficientes; es decir, en muchos casos la rotación mejorará la interpretación paliando alguna de las ambigüedades que a menudo acompañan al análisis preliminar. Las alternativas disponibles son la rotación ortogonal o la oblicua. El objetivo último de toda rotación es obtener algunos factores teóricamente significativos y, si es posible, la estructura de factores más simple. La rotación ortogonal se emplea con más frecuencia dada su presencia en todos los programas informáticos de análisis factorial, mientras que los métodos oblicuos no están tan extendidos. Además, las rotaciones ortogonales se utilizan con más frecuencia porque los procedimientos analíticos para llevar a cabo rotaciones oblicuas no están totalmente desarrollados y están todavía sujetos a controversia. Existen varias aproximaciones distintas para llevar a cabo rotaciones ortogonales u oblicuas. Sin embargo, sólo un número escaso de procedimientos de rotación oblicua está disponible en la mayoría de los programas estadísticos; por eso el investigador tendrá que aceptar probablemente alguno de los provistos.

Métodos de rotación ortogonal En la práctica, el objetivo de todos los métodos de rotación es simplificar las filas y columnas de la matriz de factores para facilitar la interpretación. En una matriz de factores, las columnas representan los factores, con cada fila correspondiendo a las cargas de las variables para cada uno de los factores. Simplificando las filas, queremos aproximar lo más posible a cero tantos valores como sea posible (es decir, maximizar la carga de una variable sobre un único factor). Simplificando las columnas, queremos aproximar lo más posible a cero tantos valores como sea posible (es decir, haciendo que el número de cargas «altas» sea el menor posible). Se han desarrollado principalmente tres aproximaciones:

QUARTIMAX El objetivo último de una rotación QUARTIMAX es simplificar las filas de una matriz de factores; esto es, QUARTIMAX se centra en rotar los factores iniciales de tal forma que una variable cargue alto sobre un factor y tan bajo como sea posible sobre los otros factores. En estas rotaciones muchas variables pueden cargar alto o cerca sobre el mismo factor por-

que la técnica se centra en las filas. El método QUARTIMAX no ha demostrado gran capacidad para generar estructuras más simples. Su dificultad está en que tiende a producir un factor general, como el primer factor, sobre el que la mayor parte, si no todas las variables, tiene cargas mayores. Con independencia del concepto que cada cual tenga de estructuras «más simples», inevitablemente se ha de tratar con agrupaciones de variables; un método que tiende a producir un factor general grande (por ejemplo, el QUARTIMAX) no responde a los objetivos de la rotación.

VARIMAX En contraste con el QUARTIMAX, el criterio VARIMAX se centra en simplificar las columnas de la matriz de factores. Con la aproximación rotacional VARIMAX, se alcanza la máxima simplificación posible si sólo hay ceros y unos en una columna.

Esto es, el método VARIMAX maximiza la suma de las varianzas de las cargas requeridas de la matriz de factores. Recuérdese que en la aproximación QUARTIMAX, muchas variables pueden cargar alto o cerca de lo alto sobre el mismo factor porque la técnica se centra en simplificar las filas. Con la aproximación rotacional VARIMAX, tiende a haber altas cargas factoriales (esto es, cercanas a –1 o +1) y algunas cargas cerca de 0 en cada columna de la matriz. Si la lógica está en que la interpretación es más fácil cuando las correlaciones variable factor están (1) cercanas a –1 o +1, indicando así una clara asociación positiva o negativa entre la variable y el factor; o (2) cercanas a 0 señalando una clara ausencia de asociación. Esta estructura resulta esencialmente sencilla. Aunque la solución QUARTIMAX es analíticamente más simple que la solución VARIMAX, VARIMAX parece suministrar una separación más clara de factores. En general, el experimento de Kaiser [21, 22] indica que el patrón factorial obtenido mediante la rotación VARIMAX tiende a resultar más robusto que el obtenido por el método QUARTIMAX cuando se analizan diferentes subconjuntos de variables. El método VARIMAX ha demostrado tener más éxito como aproximación analítica para lograr una rotación ortogonal de factores.

EQUIMAX La aproximación EQUIMAX está a medio camino entre las aproximaciones QUARTIMAX y VARIMAX. En lugar de concentrarse bien en la simplificación de las filas, bien de las columnas, procura cumplir con las dos. EQUIMAX no ha logrado una amplia aceptación y se emplea en muy raras ocasiones.

Métodos de rotación oblicua Las rotaciones oblicuas son similares a las rotaciones ortogonales, excepto que las rotaciones oblicuas permiten la existencia de factores correlacionadas en lugar de mantener la independencia entre los factores rotados. Aunque en la mayor parte de los programas estadísticos hay varias alternativas de aproximación ortogonal, suele haber escasas de rotaciones oblicuas. Por ejemplo, SPSS cuenta con OBLIMIN; SAS con PROMAX y ORTOBLIQUE; y BMDP con DQUART, DOBLIMIN y ORTOBLIQUE. Los objetivos de simplificación son comparables a los de los métodos ortogonales, con el rasgo añadido de existencia de factores correlacionados. Con esta posibilidad, el investigador ha de tener un cuidado adicional al validar los factores rotados oblicuamente, puesto que cuenta con una forma adicional (no ortogonalidad) de proceder, específica a la muestra y no generalizable, especialmente en muestras pequeñas o de bajos ratios casos/variable.

Selección del método de rotación No se han desarrollado reglas concretas que guíen al investigador en la selección de una técnica de rotación particular ortogonal u oblicua. En la mayor parte de las ocasiones, el investigador simplemente utiliza la técnica rotacional suministrada por el programa de computador. Muchos programas cuentan por defecto con la rotación VARIMAX, pero también resultan fácilmente accesibles los métodos rotacionales más comunes. Sin embargo, no existe una razón analítica incuestionable a favor de un método de rotación u otro. La elección de una rotación ortogonal u oblicua debería hacerse según las necesidades concre-

tas de un problema de investigación determinado. Si el objetivo del investigador es reducir el número de variables originales, con independencia de la significación resultante de los factores, la solución apropiada debería ser la ortogonal. Además, si el investigador desea reducir un gran número de variables a un conjunto pequeño de variables incorrelacionadas para un uso posterior en el análisis de regresión u otras técnicas de predicción, la solución ortogonal resulta la más adecuada. Sin embargo, si el objetivo último del análisis factorial es obtener varios factores teóricamente significativos, resulta apropiada una solución oblicua. Llegamos a esta conclusión dado que, realmente, muy pocos factores están incorrelacionados, como ocurre con la rotación ortogonal.

Criterios para la significación de las cargas factoriales

Al interpretar los factores, ha de adoptarse una decisión en torno a qué cargas factoriales merece la pena considerar. La siguiente exposición considera diversos aspectos relativos a la significación práctica y estadística, además de al número de variables, que afectan a la interpretación de las cargas factoriales.

Asegurar la significación práctica La primera sugerencia no está basada en afirmaciones matemáticas, sino que tiene que ver más con la significación práctica. Consiste en un tipo de regla empírica empleado frecuentemente como forma de realizar un examen preliminar de la matriz de factores. Brevemente, las cargas factoriales mayores a $\pm 0{,}30$ se considera que están en el nivel mínimo; las cargas de $\pm 0{,}40$ se consideran más importantes; y las cargas de $\pm 0{,}50$ o mayores, se consideran prácticamente significativas. Así, cuanto mayor sea el tamaño absoluto de la carga factorial, más importante resulta la carga al interpretar la matriz factorial. Dado que la carga factorial es la correlación entre la variable y el factor, el cuadrado de la carga es la cuantía de la varianza total de la variable de la que da cuenta el factor. Así, una carga de 0,30 implica aproximadamente una explicación de un 10 por ciento, y una carga de 0,50 quiere decir que un 25 por ciento de la varianza es debida al factor. Para que un factor explique un 50 por ciento de la varianza ha de contar con una carga que supere el 70 por ciento. El investigador debería darse cuenta de que cargas extremadamente elevadas (0,80 o más) no son normales y que la significación práctica es un criterio importante. Estas orientaciones son de aplicación cuando el tamaño muestral supera las 100 observaciones.

Valoración de la significación estadística Como se indicó previamente, la carga factorial representa la correlación entre la variable original y su factor. Al determinar el nivel de significación para la interpretación de las cargas, se debería emplear una aproximación similar a la utilizada para la significación estadística de los coeficientes de correlación. Sin embargo, diversas investigaciones [13] han demostrado que las cargas factoriales cuentan con errores estándar sustancialmente mayores que las correlaciones habituales, por lo que las cargas factoriales deberían evaluarse con niveles considerablemente más estrictos. El investigador puede utilizar el concepto de potencia estadística expuesto en el Capítulo 1 para especificar cargas factoriales consideradas significativas según diferentes tamaños muestrales. Con el objetivo establecido en lograr un nivel de potencia del 80 por ciento, el uso de un nivel de significación de un 0,05, y la inflación probada de los errores estándar de las cargas factoriales, la Tabla 3.2 contiene los tamaños muestrales necesarios para que cada valor de la carga factorial se considere significativo. Por ejemplo, en una muestra de 100 observaciones, las cargas factoriales de 0,55 o más son significativas. Sin embargo, en una muestra de 50, la significación implica una carga factorial de 0,75. En comparación con la anterior regla empírica que implicaba la significación para cargas del 0,30, esta aproximación consideraría a una carga de 0,30 significativa si el tamaño muestral fuera de 350 observaciones. Existen varias orientaciones prudentes cuando se compa-

TABLA 3.2. Directrices para la identificación de cargas factoriales significativas basadas en el tamaño muestral

Carga factorial	*Tamaño muestral necesario para la significación*[a]
0,30	350
0,35	250
0,40	200
0,45	150
0,50	120
0,55	100
0,60	85
0,65	70
0,70	60
0,75	50

[a] La significación se basa en un nivel de significación de 0,05 (α), un nivel de potencia del 80 por ciento y los errores estándar supuestamente dos veces mayores que los coeficientes convencionales de correlación.

Fuente: Cálculos realizados con SOLO *Power Analysis*, BMDP Statistical Software, Inc., 1993.

ran con las de la sección previa o incluso con errores estándar asociados a los coeficientes de correlación convencionales. Por ello, estas orientaciones deberían emplearse como punto de partida en la interpretación de las cargas factoriales, considerando significativas cargas factoriales bajas y de forma añadida a la interpretación basada en otras consideraciones. La siguiente sección detalla el proceso de interpretación y la función que pueden desempeñar otras consideraciones.

Ajustes basados en el número de variables Una desventaja de las aproximaciones anteriores es que no se considera el número de variables y los factores concretos que se analizan. Se ha comprobado que, a medida que el investigador se mueve del primer factor a los últimos factores, debe incrementarse el grado aceptable para considerar a una carga como significativa. El hecho de que la varianza única y la varianza del error empiecen a aparecer en los últimos factores significa que debería incluirse algún ajuste al alza en el nivel de significación [21]. Al decidir qué cargas son significativas también es importante el número de variables que se analizan. Según el número de variables analizadas, se incrementa el nivel aceptable para considerar significativa una carga que decrece. El ajuste por número de variables crece en importancia según uno se mueve del primer factor extraído a los últimos.

Resumiendo los criterios para la significación de las cargas factoriales, se pueden establecer las siguientes orientaciones: (1) a mayor tamaño muestral, menor puede ser la carga para ser considerada como significativa; (2) a mayor número de variables analizadas, menor ha de ser la carga para ser considerada como significativa; (3) a mayor número de factores, mayor ha de ser el tamaño de la carga de los últimos factores para considerarse como significativa en la interpretación.

Interpretación de la matriz de factores

La interpretación de las complejas relaciones representadas en la matriz de factores no es una tarea fácil. Siguiendo el procedimiento señalado a continuación, sin embargo, se puede simplificar considerablemente el procedimiento de interpretación factorial.

El examen de la matriz de cargas factoriales Cada columna de números en la matriz de factores representa un factor aislado. Las columnas de números son las cargas factoriales de cada variable sobre cada factor. Con el fin de identificar, el computador normalmente imprime identificando los factores de izquierda a derecha por los números 1, 2, 3, 4, etc. También las variables por su número de arriba a abajo. Para facilitar aún más la interpretación, el investigador debería escribir el nombre de cada variable en el margen izquierdo al lado del número de cada variable.

Si se ha utilizado una rotación oblicua, se presentan dos matrices de cargas factoriales. La primera es la **matriz de patrones factoriales**, que contiene las cargas que representan la contribución única de cada variable al factor. La segunda es la **matriz de estructura factorial**, que contiene las correlaciones simples entre variables y factores, pero estas cargas contienen tanto la varianza única entre variables y factores como la correlación entre factores. Según crece la correlación entre factores, es más difícil distinguir qué variables cargan únicamente sobre cada factor en la matriz de estructura factorial. Muchos investigadores suministran los resultados de la matriz de patrones factoriales.

Identificación de la mayor carga para cada variable La interpretación debería comenzar con la primera variable sobre el primer factor y moverse horizontalmente de izquierda a derecha, observando la mayor carga para cada variable sobre cada factor. Cuando se identifica la mayor carga (en valor absoluto), debe subrayarse si es significativa. Entonces la atención debe centrarse en la segunda variable, y de nuevo moviéndose de izquierda a derecha horizontalmente, comprobar la mayor carga de cada variable sobre cada factor y subrayarla. Este procedimiento debe continuar para toda variable hasta que todas las variables se hayan subrayado una vez en la mayor carga sobre un factor. Recuérdese que para tamaños muestrales menores a 100, la menor carga factorial que se considere significativa debería ser en la mayor parte de las ocasiones de $\pm 0{,}30$.

El proceso de subrayar sólo la mayor carga como significativa para cada variable es un ideal que debería perseguirse pero rara vez se consigue. Cuando cada variable tiene sólo una carga sobre un factor que es considerado significativo, la interpretación del significado de cada factor se simplifica considerablemente. En la práctica, sin embargo, muchas variables cuentan con varias cargas de tamaño moderado, todas las cuales son significativas, y el trabajo de interpretar los factores es mucho más complicado. La dificultad surge porque una variable con varias cargas significativas debe tenerse en cuenta al interpretar (etiquetar) todos los factores sobre los cuales tiene una carga significativa. Muchas soluciones factoriales no concluyen con una solución de estructura simple (una única alta carga para cada variable sólo sobre un factor). Por eso el investigador deberá continuar, tras encontrar la mayor carga para cada variable, evaluando la matriz de factores para encontrar todas las cargas significativas para una variable sobre todos los factores. Por último, el objetivo es minimizar el número de cargas significativas sobre cada fila y la matriz de factores (esto es, hacer que cada variable se asocie sólo con un factor). Una variable con varias cargas altas es candidata a ser eliminada.

Valoración de la comunalidad Una vez que todas las variables se han agrupado en sus respectivos factores, el investigador debería examinar la matriz de factores para identificar variables que no hayan sido incluidas en ningún factor. La comunalidad representa la proporción de varianza con la que contribuye cada variable a la solución final. El investigador debería observar la comunalidad de cada variable para evaluar si alcanza niveles aceptables de explicación. Por ejemplo, un investigador puede especificar que al menos sea explicada la mitad de la varianza de cada variable. Haciendo uso de este consejo, el investigador debería identificar todas las variables con comunalidades menores a 0,50 como carentes de explicación suficiente.

Si hay variables que no cargan sobre ningún factor o cuyas comunalidades se juzgan demasiado bajas, caben dos alternativas: (1) interpretar la solución tal cual es y simplemente prescindir de esas variables; o (2) evaluar cada una de esas variables para su supresión eventual. Prescindir de variables

puede resultar apropiado si el objetivo es únicamente la reducción de datos, pero el investigador todavía debe percatarse de que las variables en cuestión están pobremente representadas en la solución factorial. La consideración sobre su eliminación debe depender de la contribución conjunta de las variables para el investigador además del índice de comunalidad. Si la variable en cuestión es de escasa importancia para el objetivo del estudio o cuenta con una comunalidad inaceptable, podría ser eliminada y se procedería posteriormente a especificar el modelo factorial excluyendo esa variable.

Etiquetación de los factores Cuando se ha obtenido una solución factorial en que todas las variables tienen una carga significativa sobre un factor, el investigador procura atribuir un significado al patrón de cargas factoriales. Las variables con mayores cargas se consideran más importantes y tienen mayor influencia sobre el nombre o etiqueta seleccionada para representar al factor. Por eso, el investigador examinará todas las variables agrupadas en un factor particular y, poniendo mayor énfasis en las variables con mayor carga factorial, intentará asignar un nombre o etiqueta al factor que refleje con precisión las variables cargadas sobre el factor. Los signos se interpretan como otros coeficientes de correlación. Sobre cada factor, signos iguales significan que las variables están positivamente relacionadas, signos opuestos significan que las variables están negativamente relacionadas. En soluciones ortogonales los factores son independientes unos de otros. Por tanto, los signos de las cargas factoriales se relacionan sólo con el factor en el cual aparecen, no con otros factores en la solución.

Esta etiqueta no viene asignada por el análisis factorial realizado por el programa de computador; en su lugar, la etiqueta se fabrica intuitivamente de acuerdo a la conveniencia para representar las dimensiones subyacentes de un factor concreto. El resultado final será el nombre o etiqueta que representa cada uno de los factores obtenidos con tanta precisión como sea posible.

En algunas ocasiones, no es posible asignar un nombre a cada uno de los factores. Cuando surge tal situación, el investigador desearía etiquetar un factor o factores derivados de la solución como «indefinidos». En tales casos el investigador interpreta sólo aquellos factores que son significativos y elude aquellos indefinidos o menos significativos. Al describir la solución factorial, el investigador indica que esos factores se obtuvieron pero que eran indefinidos y que sólo se interpretan aquellos factores que representan relaciones significativas.

Como se expuso anteriormente, la selección de un número concreto de factores y el método de rotación están interrelacionados. Se pueden llevar a cabo varias rotaciones adicionales de prueba y comparando la interpretación factorial para las diferentes rotaciones ensayadas, el investigador puede seleccionar el número de factores a extraer. En resumen, la habilidad para asignar algún significado a los factores, o para interpretar la naturaleza de las variables, son consideraciones extremadamente importantes para determinar el número de factores a extraer.

Sexto paso: Validación del análisis factorial

La sexta etapa comprende la evaluación del grado de generabilidad de los resultados para la población y la influencia potencial de causas o individuos sobre los resultados globales. El aspecto de la generabilidad es esencial en cada uno de los métodos multivariantes, pero es especialmente relevante en los métodos de interdependencia por describir una estructura de datos que debería ser representativa también de la población. El método más directo de validación de los resultados consiste en adoptar una perspectiva de confirmación, valorando la replicabilidad de los resultados, bien dividiendo la muestra con los datos originales, bien con una muestra adicional. La comparación de los resultados de dos o más modelos factoriales siempre ha sido problemática. Sin embargo, existen varias alternativas para realizar una comparación objetiva. El auge del análisis factorial confirmatorio (CFA) a través de la modelización de ecuaciones estructurales supone una alternativa, pero generalmente es más complicado y requiere software adicional como LISREL o EQS [4, 20]. En el Capítulo 11, se discute el análisis factorial confirmatorio en detalle. Además del CFA, se han propuesto otros métodos que van desde un simple índice de adecuación [10] a programas

(FMATCH) diseñados especialmente para valorar la correspondencia entre matrices de factores [34]. Estos métodos cuentan con un uso ocasional, debido en parte a (1) la percepción de ausencia de sofisticación y (2) la no disponibilidad de software o programas analíticos que automaticen las comparaciones. Por eso, cuando CFA no es apropiado estos métodos facilitan una base objetiva para la comparación.

Otro aspecto de la generalización es la estabilidad de los resultados del modelo factorial. La estabilidad factorial depende primeramente del tamaño muestral y del número de casos por variable. El investigador siempre está obsesionado por contar con el mayor tamaño muestral posible y desarrollar modelos parsimoniosos que incrementen la ratio-casos-por-variable. Si el tamaño muestral lo permite, el investigador puede dividir aleatoriamente la muestra en dos subconjuntos y estimar los modelos factoriales de cada uno. La comparación de las dos matrices factoriales resultantes suministrará una valoración de la robustez de la solución a lo largo de la muestra.

Además de la generabilidad, otro aspecto de importancia para la validación del análisis factorial es la detección de observaciones influyentes. Las discusiones del Capítulo 2 sobre la detección de atípicos y del Capítulo 4 sobre observaciones influyentes en la regresión se deben aplicar también al análisis factorial. El investigador procura estimar el modelo con y sin observaciones identificadas como atípicas para valorar su influencia sobre los resultados. También, como se discutió en el Capítulo 4, son de aplicación al análisis factorial varias medidas de influencia que reflejan la posición relativa de una observación respecto a las otras (por ejemplo, el ratio de la covarianza). Finalmente, se han propuesto métodos específicos de análisis factorial para identificar observaciones influyentes [11], pero su complejidad ha restringido su aplicación.

Séptimo paso: Usos adicionales de los resultados del análisis factorial

Dependiendo de los objetivos de partida al aplicar el análisis factorial, el investigador puede detenerse en la interpretación de los factores o proseguir con uno de los métodos de reducción de datos. Si el objetivo simplemente consiste en identificar combinaciones lógicas de variables y una mejor comprensión de las relaciones entre variables, entonces basta con la interpretación de los factores. Ésta facilita una base empírica para enjuiciar la estructura de las variables y la influencia de esta estructura cuando se interpretan los resultados a partir de otras técnicas multivariantes. Si el objetivo, sin embargo, es identificar variables apropiadas para aplicaciones subsiguientes de otras técnicas estadísticas, se empleará alguna forma de reducción de datos. Las alternativas consisten en (1) examinar la matriz de factores y seleccionar la variable con mayor carga factorial como un representante de una dimensión factorial particular, o (2) reemplazar el conjunto original de variables por uno totalmente nuevo con menos variables creado a partir de escalas aditivas o de la puntuación de factores. Cada alternativa suministrará nuevas variables para ser usadas, por ejemplo, como variables independientes en una regresión o en el análisis discriminante, o como variables dependientes en el análisis multivariante de la varianza, o incluso como las variables agrupadas en el análisis cluster. Cada una de estas alternativas para la reducción de datos se exponen en las siguientes secciones.

Selección de variables suplentes para el análisis subsiguiente

Si el objetivo del investigador es sencillamente identificar variables apropiadas para la aplicación subsiguiente de otras técnicas estadísticas, cuenta con la alternativa de examinar la matriz factorial y seleccionar la variable con la mayor carga factorial sobre cada factor para que actúe como **variable suplente** del factor. Este enfoque es simple y directo sólo cuando una variable tiene una

carga factorial que es sustancialmente mayor que las otras cargas factoriales. En muchas ocasiones, sin embargo, el proceso de selección es mucho más difícil porque dos o más variables tienen cargas que son significativas y bastante cercanas entre sí. Estos casos requieren un examen crítico de las cargas factoriales de tamaño aproximado y sólo una como representativa de una dimensión concreta. Esta decisión debería basarse en el conocimiento previo de la teoría por parte del investigador que pueda sugerir que una variable con preferencia a otra pueda ser representativa de la dimensión. Además, el investigador puede contar con un conocimiento que le sugiera que una variable con una carga ligeramente inferior es de hecho más fiable que la variable con la mayor carga. En tales casos, puede elegir la variable con carga ligeramente inferior como la mejor variable suplente de un factor concreto.

La aproximación de seleccionar una única variable como representativa del factor – aunque simple y manteniendo la variable original – cuenta con varios inconvenientes potenciales. En primer lugar, no orienta sobre el error de medida que aparece cuando se emplean medidas únicas (véase la siguiente sección para una discusión más detallada) y se corre, además, el riesgo de confundir potencialmente los resultados seleccionando sólo una única variable para representar un resultado posiblemente más complejo. Por ejemplo, supongamos que las variables que representan competitividad en precio, calidad del producto y valor, cargan en varios factores. La selección de cualquiera de estas variables aislada daría lugar a interpretaciones sustancialmente distintas en cualquier análisis subsiguiente, aunque las tres pueden estar tan próximamente relacionadas como para excluir tal acción. En segundo lugar, en casos donde varias cargas elevadas complican la selección de una única variable, el investigador puede no tener otra elección que la de emplear el análisis factorial como la base para calcular una escala aditiva o la puntuación de factores para usar como variables suplentes. El objetivo, como en el caso de seleccionar una única variable, es representar de la mejor forma posible la naturaleza básica del factor o componente.

Creación de escalas aditivas

En el Capítulo 1 se introdujo el concepto de escala aditiva, que está formada por la combinación de varias variables individuales dentro de una única **medida compuesta**. Simplificando, se combinan todas las variables que cargan alto sobre un factor, y el total – o más normalmente la puntuación media de las variables – se emplea como variable de sustitución. Una escala aditiva cuenta con dos ventajas concretas. Primeramente, es una forma de eludir en alguna forma el error de medida inherente a todas las variables observadas. El **error de medida** es el grado en el cual los valores observados no son representativos de los valores «verdaderos» debido a cierto número de razones, desde errores reales (por ejemplo, errores en la entrada de los datos) a la incapacidad de los individuos de suministrar información con precisión. El error de medida enmascara cualquier relación (por ejemplo, correlaciones o comparación de medias de grupos) y hace más difícil la estimación en los modelos multivariantes. La escala aditiva reduce el error de medida utilizando **indicadores** múltiples (variables) para reducir la desconfianza sobre una única respuesta. Empleando la «media» o la respuesta «típica» a un conjunto de variables relacionadas, el error de medida que podría tener lugar en una única respuesta se reducirá.

Una segunda ventaja de la escala aditiva es su capacidad para representar los múltiples aspectos de un concepto en una única medida. En muchas ocasiones empleamos más variables en nuestros modelos multivariantes en un intento de representar las muchas «facetas» de un concepto que sabemos es bastante complejo. Pero al actuar así, complicamos la interpretación de los resultados debido a la redundancia de la información asociada con el concepto. Por eso, nos gustaría no sólo obtener una descripción mejor de los conceptos utilizando múltiples variables, sino también mantener la parsimonia en el número de variables de nuestros modelos multivariantes. La escala aditiva, cuando se construye apropiadamente, combina los múltiples indicadores en una medida única representando lo que se mantiene en común a lo largo del conjunto de medidas.

El proceso de construcción de la escala está fundamentado teórica y empíricamente en una serie de disciplinas que incluyen la teoría psicométrica, la sociología y el marketing. Aunque un tra-

tamiento completo de las técnicas y aspectos involucrados están más allá del alcance de este libro, existen fuentes excelentes para un estudio más extenso de estas materias [2, 12, 19, 29, 30]. Adicionalmente hay una serie de compilaciones de escalas existentes que pueden aplicarse en varias situaciones [3, 7, 31]. Aquí expondremos, sin embargo, cuatro aspectos básicos en la construcción de cualquier escala aditiva: la definición conceptual, la dimensionalidad, la fiabilidad y la validación.

Definición conceptual El punto de partida para construir una escala aditiva es su **definición conceptual**. La definición conceptual especifica las bases teóricas de la escala aditiva definiendo el concepto que está representado en términos aplicables al contexto de investigación. En la investigación académica, las definiciones teóricas están basadas en investigación previa que define el carácter y naturaleza de un concepto. En un ámbito de gestión empresarial, los conceptos concretos pueden definirse con relación a los objetivos propuestos, tales como la imagen, el valor o la satisfacción. En cualquier caso, la definición conceptual es la que orienta y concreta el carácter y tipo de ítems que son candidatos a ser incluidos en la escala.

La **validación del contenido** es la evaluación de la correspondencia de las variables incluidas en la escala aditiva con su definición conceptual. Esta forma de validación, también conocida como **validación aparente**, sirve para apreciar subjetivamente la correspondencia entre los ítems individuales y el concepto a través de evaluaciones de expertos, contrastes previos con múltiples subpoblaciones, u otros medios. El objetivo es asegurar que los ítems de la escala abarquen algo más que aspectos empíricos, también consideraciones teóricas y prácticas [12, 30].

Dimensionalidad Un supuesto subyacente y requisito esencial para construir una escala aditiva es que los ítems sean unidimensionales, lo que significa que estén fuertemente asociados unos con otros representando un único concepto [19, 23]. El análisis factorial sirve de apoyo realizando una valoración empírica de la dimensionalidad del conjunto de ítems determinando el número de factores y las cargas de cada variable sobre el factor o factores. El contraste de unidimensionalidad consiste en que en una escala aditiva los ítems carguen de forma elevada en un único factor [1, 19, 23, 27]. Si se propone que una escala aditiva cuente con múltiples dimensiones, cada dimensión reflejará un factor aislado. El investigador puede evaluar la unidimensionalidad bien con un análisis factorial exploratorio, como se discutió en este capítulo, o bien un análisis factorial confirmatorio, como se describe en el Capítulo 11.

Fiabilidad La fiabilidad es el grado de consistencia entre las múltiples medidas de una variable. Una forma de fiabilidad es el test-retest por el cual la consistencia se mide entre las respuestas de un individuo en dos momentos del tiempo. El objetivo es asegurar que las respuestas no varían demasiado a lo largo de períodos temporales por lo que una medida tomada en cualquier momento del tiempo es certera. Una segunda medida de la fiabilidad más utilizada es la consistencia interna que se aplica a la consistencia entre las variables en una escala aditiva. La motivación para la consistencia interna es que los ítems individuales o indicadores de la escala deberían estar midiendo las mismas construcciones y, de esta forma, estar altamente intercorrelacionadas [12, 27].

Debido a que no hay un único ítem que sea una medida perfecta de un concepto, debemos delegar en un conjunto de medidas de diagnosis para valorar la consistencia interna. En primer lugar, existen varias medidas que se relacionan con cada ítem aislado, incluyendo la correlación ítem-total (la correlación del ítem con la puntuación de la escala aditiva) o la correlación inter-ítem (la correlación entre ítems). La práctica empírica sugiere que la correlación ítem-total exceda de 0,50 y que las correlaciones inter-ítem excedan de 0,30 [30]. El segundo tipo de medida de diagnosis es el coeficiente de fiabilidad que valora la consistencia de la escala entera, el **alfa de Cronbach** [27, 28] que es la medida más extensamente utilizada. El acuerdo general sobre el límite inferior para al alfa de Cronbach es de 0,70 [30, 31], aunque puede bajar a 0,60 en la investiga-

ción exploratoria [30]. Un aspecto en la valoración del alfa de Cronbach es su relación positiva con el número de ítems de la escala. Debido a que al aumentar el número de ítems, incluso con el mismo grado de intercorrelación, se incrementará el valor de la fiabilidad, los investigadores deben imponer requisitos más restrictivos para escalas con un mayor número de ítems. También están disponibles medidas de fiabilidad derivadas del análisis factorial confirmatorio. Dentro de ellas están la fiabilidad compuesta y la varianza media extraída, ambas discutidas con mayor detalle en el Capítulo 11.

Cada uno de los programas estadísticos más utilizados contiene ahora módulos de evaluación de la fiabilidad, de tal forma que el investigador está provisto de un análisis completo de medidas tanto específicas de cada ítem como medidas de fiabilidad globales. En toda escala aditiva debería analizarse la fiabilidad para asegurar su adecuación antes de proceder a una evaluación de su validación.

Validación Habiendo asegurado que una escala (1) se adecúa a su definición conceptual, (2) es unidimensional, y (3) alcanza los niveles necesarios de fiabilidad, el investigador debe realizar una evaluación final: la validación de la escala. La **validación** es la medida en que una escala o un conjunto de medidas representa con precisión el concepto de interés. Ya hemos discutido una forma de validación —el contenido o validación por confrontación— en relación a las definiciones conceptuales. Otras formas de validación se miden empíricamente por la correlación entre los conjuntos de variables definidas teóricamente. Las tres formas más extensamente admitidas de validación son la convergente, la discriminante y la nomológica [8, 29]. La validación convergente valora el grado en el cual dos medidas del mismo concepto están correlacionadas. Aquí el investigador debe buscar medidas alternativas de un concepto y correlacionarlas con la escala aditiva. Altas correlaciones indican que la escala está midiendo el concepto que se pretendía. La validación discriminante es el grado en el cual dos conceptos conceptualmente parecidos difieren. El contraste empírico de nuevo es la correlación entre las medidas, pero esta vez la escala aditiva está correlacionada con una medida parecida, pero conceptualmente distinta. Ahora la correlación debería ser baja, demostrando que la escala aditiva es conceptualmente distinta de otro concepto parecido. Finalmente, la validación nomológica refleja el grado en que la escala aditiva hace predicciones precisas de otros conceptos en un modelo basado en la teoría. El investigador debe identificar relaciones apoyadas en investigación previa o en principios aceptados y evaluar si la escala cuenta con las correspondientes relaciones. En resumen, la validación convergente confirma que la escala está correlacionada con otras medidas conocidas del concepto, la validación discriminante asegura que la escala es suficientemente distinta de otros conceptos parecidos que sean distintos, y la validación nomológica determina si la escala demuestra las relaciones cuya existencia se deriva de la teoría y/o de investigación previa.

Existen varios métodos para evaluar la validación, que van desde el multirasgo, las matrices multimétodo (MTMM) a las aproximaciones basadas en ecuaciones estructurales. Aunque vayan más allá del alcance de este libro, existen una serie de fuentes que orientan sobre el conjunto de métodos disponibles y de los aspectos implicados por las técnicas concretas [8, 20, 29].

Resumen

Las escalas aditivas, uno de los desarrollos recientes en la investigación académica, cuentan con aplicación creciente en investigación aplicada y también en gestión empresarial. La capacidad de la escala aditiva para plasmar conceptos complejos en una única medida reduciendo el error de medida supone un añadido valioso en cualquier análisis multivariante. El análisis factorial ofrece al investigador una evaluación empírica de las relaciones entre las variables, esencial en la formación de los fundamentos conceptuales y empíricos de la escala aditiva por medio de la evaluación de la validación del contenido y la dimensionalidad de la escala.

Cálculo de la puntuación factorial

La tercera alternativa para crear un conjunto más pequeño de variables es reemplazar el conjunto original por medio del cálculo de la puntuación factorial. Las **puntuaciones factoriales** constituyen también medidas compuestas de cada factor calculadas para cada sujeto. Conceptualmente, la puntuación factorial representa el grado en el cual cada individuo puntúa en el grupo de ítems que cuentan con cargas elevadas sobre un factor. Por tanto, mayores valores en las variables con altas cargas sobre un factor resultarán en una mayor puntuación factorial. Una de las características clave que diferencia la puntuación factorial de la escala aditiva es que la puntuación factorial se calcula en base a las cargas factoriales, mientras que la escala aditiva se calcula combinando sólo las variables elegidas. En consecuencia, aunque el investigador sea capaz de caracterizar un factor por medio de las variables con cargas mayores, se debe también atender a las cargas de las otras variables, aunque menores, y su influencia sobre la puntuación factorial.

La mayoría de los programas estadísticos puede calcular puntuaciones factoriales fácilmente para cada elemento muestral. Seleccionando la alternativa de puntuación factorial, se graban estas puntuaciones para su uso en análisis subsiguientes. Una desventaja de la puntuación factorial es que no se pueden replicar fácilmente en otros estudios debido a que están basados en la matriz factorial obtenida en cada estudio. La replicación de la misma matriz factorial en distintos estudios requiere un importante trabajo de programación.

Selección entre los tres métodos

Para elegir entre las tres alternativas para reducir datos el investigador debe adoptar una serie de decisiones. La primera elección está en seleccionar una única variable suplente para cada factor o calcular una medida compuesta. La variable suplente única cuenta con las ventajas de ser de sencillo tratamiento e interpretación, pero tiene las desventajas de no representar las otras «facetas» del factor y su propensión al error de medida. Si el investigador desea emplear alguna forma de medida compuesta, debe elegir entre la puntuación factorial y las escalas aditivas. Ambas cuentan con ventajas y desventajas y no hay un criterio tajante aplicable a todas las situaciones. La puntuación factorial tiene la ventaja de representar un compuesto de las cargas de las variables sobre el factor, aunque esto supone también una desventaja potencial al contar todas las variables con algún grado de influencia en el cálculo de la puntuación factorial y hacer la interpretación más difícil. La escala aditiva está a medio camino entre la variable suplente y la alternativa de la puntuación factorial. Es una medida compuesta, como la puntuación factorial, reduciendo por tanto el error de medida y representando múltiples facetas de un concepto. Aunque similar a la aproximación de variable suplente incluye sólo las variables que cargan alto sobre el factor y excluye aquellas con escasos efectos. Además, su fácil replicación entre muestras es similar al enfoque de variable suplente. Finalmente, como las variables suplentes, las escalas aditivas no son necesariamente ortogonales, mientras que los factores pueden ser ortogonales o incorrelados, si se necesita evitar complicaciones en su uso en otras técnicas multivariantes. La regla de decisión, por tanto, sería que si los datos se emplean sólo en la muestra original o se debe mantener la ortogonalidad, la puntuación factorial es la adecuada. Si se desea la transferibilidad o la generalización, las escalas aditivas o las variables suplentes son más apropiadas. Si la escala aditiva está bien construida, validada y es fiable, es probable que sea la mejor alternativa. Pero si la escala aditiva no está contrastada y revisada, con poca o ninguna prueba de fiabilidad o validación, deberían considerarse en su lugar las variables suplentes si no es posible un análisis añadido que mejore la escala aditiva.

Un ejemplo ilustrativo

En las secciones precedentes se han discutido las principales preguntas concernientes a la aplicación del análisis factorial en el contexto de la construcción de modelos introducido en el Capítulo 1. Para aclarar más estos temas, exponemos un ejemplo ilustrativo de aplicación del análisis

factorial basado en ítems de la base de datos presentada en el Capítulo 1. Nuestra discusión del ejemplo empírico sigue también el proceso de construcción de modelos en seis pasos. Se presentan en primer lugar los tres primeros, comunes al análisis factorial común y al de componentes principales. Posteriormente, los pasos cuarto a sexto se dedican al análisis de componentes principales, junto a los ejemplos de los usos adicionales del análisis factorial. Concluimos con un análisis de las diferencias para el análisis factorial común en los pasos cuarto y quinto.

Primer paso: Objetivos del análisis factorial

El análisis factorial puede identificar la estructura de un conjunto de variables además de suministrar un proceso para la reducción de datos. En nuestro ejemplo, se examinan las percepciones de HATCO sobre siete atributos (X_1 a X_7) para (1) comprender si esas percepciones pueden ser «agrupadas» y (2) reducir las siete variables a un número menor. Incluso el número relativamente menor de percepciones examinadas presenta aquí un cuadro complejo de 21 correlaciones aisladas. Agrupando las percepciones, HATCO dispondrá de un panorama que le permitirá comprender a sus clientes y lo que éstos piensan sobre HATCO. Si las siete variables pueden representarse en un número menor de variables compuestas, las otras técnicas multivariantes pueden ser más parsimoniosas. Por supuesto, este enfoque supone que existe un determinado orden subyacente en los datos analizados.

Segundo paso: Diseño del análisis factorial

La comprensión de la estructura de las percepciones de las variables requiere un análisis factorial de tipo R y una matriz de correlación entre las variables, no entre las respuestas. Todas las variables son métricas y forman un conjunto homogéneo apropiado para el análisis factorial. Atendiendo a la adecuación del tamaño muestral en este ejemplo hay un ratio de observaciones a variables de 14 a 1, que entra en los límites aceptables. Además, el tamaño muestral de 100 supone una base adecuada para el cálculo de las correlaciones entre variables.

Tercer paso: Supuestos de análisis factorial

Los supuestos estadísticos subyacentes afectan al análisis factorial en la medida en que afectan a las correlaciones obtenidas. Incumplimientos de la normalidad, la homocedasticidad y la linealidad pueden reducir las correlaciones entre variables. Estos supuestos se examinan en el Capítulo 2. Se invita al lector a la revisión de los resultados.

El investigador puede también evaluar la aplicación del factorial con el análisis de la matriz de correlaciones. El primer paso es el examen visual de las correlaciones, identificando aquellas que sean estadísticamente significativas. La Tabla 3.3 muestra la matriz de correlaciones para las siete percepciones de HATCO. La inspección de la matriz de correlaciones revela que 11 de las 21 correlaciones (57 por ciento) son significativas al nivel del 0,01.

Este análisis suministra una base adecuada para el siguiente nivel, el examen empírico de la suficiencia del análisis factorial tanto sobre bases globales como para cada variable.

El siguiente paso es valorar la significación de la matriz de correlación con el contraste de Barlett. En este ejemplo, las correlaciones, cuando se toman conjuntamente, son significativas a un nivel del 0,0001 (véase Tabla 3.3). Pero con este test se contrasta sólo la presencia de correlaciones no nulas, no el patrón de esas correlaciones. El otro contraste global es la medida de adecuación muestral (MSA), que en este caso cae en el rango de no aceptación (por debajo del 0,50) con un valor de 0,446. El examen de los valores para cada variable identifica tres variables (X_1, X_2 y X_5) que también tienen valores por debajo de 0,50. Debido a que X_5 tiene el menor valor MSA, será omitida con el fin de obtener un conjunto de variables que puedan exceder los niveles mínimos aceptables MSA.

TABLA 3.3. Evaluación de la conveniencia del análisis factorial: Correlaciones, medidas de suficiencia de muestreo y correlaciones parciales entre variables

	Correlación entre las variables						
Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1 Velocidad de entrega	1,00	–0,35*	0,51*	0,05	0,61*	0,08	–0,48*
X_2 Nivel de precio		1,00	–0,49*	0,27*	0,51*	0,19	0,47*
X_3 Flexibilidad de precio			1,00	–0,12	0,07	–0,03	–0,45*
X_4 Imagen del fabricante				1,00	0,30*	0,79*	0,20
X_5 Servicio					1,00	0,24*	–0,06
X_6 Imagen del personal de ventas						1,00	0,18
X_7 Calidad de producto							1,00

* Indica las correlaciones significativas al nivel 0,01.

Medida de suficiencia de muestreo general: 0,446
Contraste de esfericidad de Bartlett: 567,5 Significación: 0,0000

	*Medidas de suficiencia de muestreo y correlaciones parciales**						
Variable	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1 Velocidad de entrega	0,344						
X_2 Nivel de precio	0,957	0,330					
X_3 Flexibilidad de precio	0,018	0,155	0,913				
X_4 Imagen del fabricante	0,149	0,134	0,095	0,558			
X_5 Servicio	–0,978	–0,975	–0,091	–0,173	0,288		
X_6 Imagen del personal de ventas	–0,060	–0,045	–0,085	–0,766	0,052	0,552	
X_7 Calidad de producto	–0,016	–0,141	0,140	–0,039	0,088	–0,092	0,927

* Los valores de la diagonal principal son medidas de suficiencia de muestreo para variables individuales; los valores fuera de la diagonal son correlaciones anti-imagen (correlaciones parciales negativas).

La Tabla 3.4 contiene la matriz de correlación para el conjunto revisado de variables (X_1, X_2, X_3, X_4, X_6 y X_7) junto a las medidas de adecuación muestral y el valor del contraste de Barlett. En la matriz de correlaciones reducida, 7 de las 15 correlaciones son estadísticamente significativas. Como con el conjunto completo de variables, el contraste de Barlett muestra que las correlaciones no nulas existen a un nivel de significación del 0,0001. El conjunto reducido de variables colectivamente alcanza el umbral necesario de suficiencia muestral con un valor MSA de 0,665. Cada una de las variables también excede el valor del umbral, indicando que el conjunto reducido de variables alcanza los requisitos fundamentales para el análisis factorial. Finalmente, con la excepción de una correlación parcial (X_4 y X_6), las demás son todas bastante bajas, lo que es otro indicador de la fortaleza de las relaciones entre las variables del conjunto reducido. Todas estas medidas indican que el conjunto reducido de variables es apropiado para el análisis factorial, y que se puede proseguir con los siguientes pasos.

Análisis de factores componentes: Del paso cuarto al séptimo

Como se indicó anteriormente, los procedimientos de análisis factorial están basados en el cálculo inicial de una tabla completa de intercorrelaciones entre las variables (matriz de correlación). Esta matriz de correlación se transforma entonces por medio de la estimación de un modelo de factores para obtener la matriz de factores. Las cargas de cada variable sobre los factores se interpretan entonces para identificar la estructura subyacente de las variables, en este caso percepciones de HATCO. Estos pasos del análisis factorial, contenidos en los pasos cuarto a siete, se examinan pri-

TABLA 3.4. Evaluación de la conveniencia del análisis factorial para la serie de variables revisada: correlaciones, medidas de suficiencia de muestreo y correlaciones parciales entre variables

	Correlación entre las variables					
Variable	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	1,00	–0,35*	0,51*	0,05	0,08	–0,48*
X_2 Nivel de precio		1,00	–0,49*	0,27*	0,19	0,47*
X_3 Flexibilidad de precio			1,00	–0,12	–0,03	–0,45*
X_4 Imagen del fabricante				1,00	0,79*	0,20
X_6 Imagen del personal de ventas					1,00	0,18
X_7 Calidad de producto						1,00

* Indica las correlaciones significativas al nivel 0,01.

Medida de suficiencia de muestreo general: 0,665
Contraste de esfericidad de Bartlett: 205,965 Significación: 0,0000

	*Medidas de suficiencia de muestreo y correlaciones parciales**					
Variable	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	0,721					
X_2 Nivel de precio	0,074	0,787				
X_3 Flexibilidad de precio	–0,338	0,301	0,748			
X_4 Imagen del fabricante	–0,098	–0,160	0,081	0,542		
X_6 Imagen del personal de ventas	–0,045	0,026	–0,081	–0,769	0,532	
X_7 Calidad de producto	0,331	–0,253	0,149	–0,024	–0,097	0,779

* Los valores de la diagonal principal son medidas de suficiencia de muestreo para variables individuales los valores fuera de la diagonal son correlaciones anti-imagen (correlaciones parciales negativas).

mero para el análisis de componentes principales. Entonces, se lleva a cabo un análisis factorial común y se comparan los dos modelos de factores.

Cuarto paso: *Derivación de los factores y valoración del ajuste global*

El primer paso es seleccionar el número de componentes que se van a mantener para un análisis posterior. La Tabla 3.5 contiene la información relativa a los seis posibles factores y su poder explicativo relativo expresado por sus autovalores. Además de valorar la importancia de cada componente, podemos emplear los autovalores como ayuda para seleccionar el número de factores. Si aplicamos el criterio de raíz latente, se mantendrán dos componentes. El contraste de caída (Figura 3.9), sin embargo indica que los tres factores pueden resultar apropiados. Al mirar el

TABLA 3.5. Los resultados para la extracción de los factores de componente

Factor	*Autovalor*	*Porcentaje de varianza*	*Porcentaje de varianza acumulado*
1	2,51349	41,9	41,9
2	1,73952	29,0	70,9
3	0,59749	10,0	80,8
4	0,52956	8,8	89,7
5	0,41573	6,9	96,6
6	0,20422	3,4	100,0

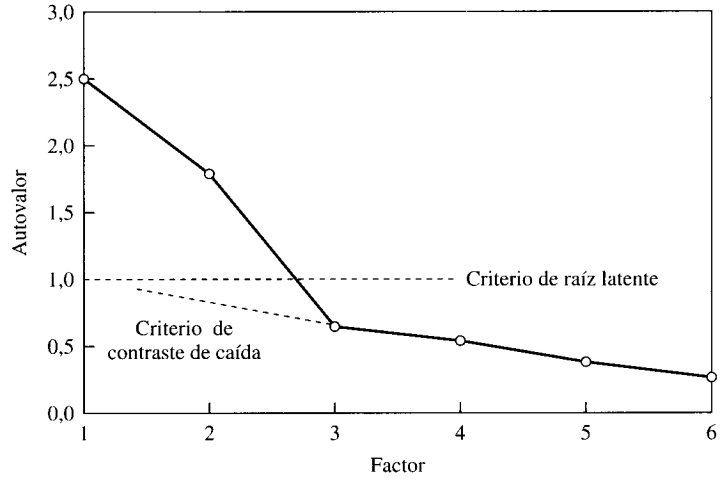


FIGURA 3.9. Contraste de caída para el análisis de componente.

autovalor para el tercer factor, se determinó que su valor bajo (0,597) relativo al criterio de raíz latente de 1,0 impide su inclusión. Si su autovalor hubiera estado cercano a 1, entonces también podría haberse considerado su inclusión. Estos resultados ilustran la necesidad de criterios de decisión múltiples al decidir el número de componentes que se mantienen. Los dos factores mantenidos representan el 71 por ciento de la varianza de las seis variables.

Quinto paso: *Interpretación de los factores*

El resultado del cuarto paso se presenta en la Tabla 3.6, la matriz de factores del análisis de componentes no rotados. Para empezar el análisis, expliquemos en primer lugar las cantidades incluidas en la tabla. Se muestran tres columnas de números. Las dos primeras son los resultados para dos factores que van a extraerse (esto es, cargas factoriales sobre cada variable para cada factor). La tercera columna suministra estadísticas de resumen detallando cómo cada variable está «explicada» por los dos componentes, que son discutidos en la siguiente sección. La primera fila de números al final de cada columna es la suma de los cuadrados de las cargas factoriales (autovalores) e indica la importancia relativa de cada factor dando cuenta de la varianza asociada con el conjunto de variables que se analiza. Nótese que la suma de los cuadrados para los dos factores son

TABLA 3.6. Matriz de factores de análisis de componente no rotado

	Factores		
Variables	1	2	*Comunalidades*
X_1 Velocidad de entrega	–0,627	0,514	0,66
X_2 Nivel de precio	0,759	–0,068	0,58
X_3 Flexibilidad de precio	–0,730	0,337	0,65
X_4 Imagen del fabricante	0,494	0,798	0,88
X_6 Imagen del personal de ventas	0,425	0,832	0,87
X_7 Calidad de producto	0,767	–0,168	0,62
			Total
Suma de cuadrados (autovalor)	2,51	1,74	4,25
Porcentaje de traza*	41,9	29,0	70,9

* Traza = 6,0 (suma de los autovalores).

2,51 y 1,74 respectivamente. Como se esperaba, la solución factorial ha extraído los factores en el orden de su importancia, con el factor 1 dando cuenta de la mayor parte de varianza y el 2 algo menos. En el extremo derecho de la fila está el número 4,25, que representa la suma explicada total de los cuadrados (2,51 ± 1,74). La suma total de los factores al cuadrado representa la cantidad total de varianza explicada por la solución factorial.

La cantidad total de varianza explicada por la solución factorial (4,25) puede compararse con la variación total en el conjunto de variables tal como viene representada por la traza de la matriz factorial. La traza es la varianza total a explicar y es igual a la suma de los autovalores del conjunto de variables. En el análisis de componentes principales, la traza es igual al número de variables puesto que cada variable tiene un posible autovalor igual a 1,0. Los porcentajes de traza explicados por cada uno de los dos factores (41,9 por ciento y 29,0 por ciento respectivamente) aparecen como la última fila de valores de la Tabla 3.6. El porcentaje de traza se obtiene dividiendo la suma de los cuadrados de cada factor por la traza para el conjunto de variables analizado. Por ejemplo, dividiendo la suma de los cuadrados de 2,51 para el factor 1 por la traza de 6,0 resulta un porcentaje de traza del 41,9 por ciento para el factor 1. Sumando los porcentajes de traza para cada uno de los dos factores, obtenemos el porcentaje total de traza extraído para la solución factorial, que puede emplearse como índice para determinar con qué bondad da cuenta una solución factorial concreta de lo que todas las variables representan conjuntamente. Si todas las variables son muy diferentes unas de otras, este índice será bajo. Si las variables pertenecen a uno o más grupos altamente redundantes o relacionados, o si los factores extraídos dan cuenta de todos los grupos, el índice se aproximará al 100 por cien. El índice para la solución presente muestra que el 70,9 por ciento del total de la varianza está representado por la información contenida en la matriz factorial de la solución para los dos factores. Por tanto, el índice para esta solución es elevado, y las variables están de hecho altamente relacionadas unas con otras.

La fila suma de las cargas de los factores al cuadrado se muestra en el extremo derecho de la Tabla 3.6. Estas cifras, señaladas en la tabla como comunalidades, muestran la cantidad de varianza en una variable que es explicada por los dos factores tomados conjuntamente. El tamaño de la comunalidad es un índice útil para valorar cuánta varianza en una variable concreta viene explicada por la solución factorial. Grandes comunalidades indican que la solución factorial ha extraído una gran cantidad de varianza en una variable. Pequeñas comunalidades muestran que una porción sustancial de la varianza en una variable viene explicada por los factores. Por ejemplo, la comunalidad de 0,65 para la variable X_3 indica que tiene menos en común con las otras variables incluidas en el análisis de lo que lo hace la variable X_4, que tiene una comunalidad de 0,88. Ambas variables, sin embargo, todavía «comparten» en torno a un medio de su varianza con los otros factores. Habiendo definido los diversos elementos de la matriz de factores no rotados, examinemos los patrones de cargas factoriales. Como se anticipó, el primer factor da cuenta de la mayor cantidad de varianza y es un factor general, en que todas las variables tienen una carga alta. Las cargas del segundo factor muestran tres variables (X_1, X_4 y X_6) que también tienen cargas altas. Basados en este patrón de altas cargas factoriales, la interpretación resultaría bastante difícil y teóricamente menos significativa. Por tanto, el investigador debería proceder a rotar la matriz factorial para redistribuir la varianza de los primeros factores a los últimos. De la rotación debe resultar un patrón factorial más sencillo y teóricamente más significativo.

La aplicación de la rotación ortogonal (VARIMAX) La matriz de factores del análisis de componentes rotados VARIMAX se muestra en la Tabla 3.7. Nótese que la cantidad total de varianza es la misma en la solución rotada de lo que lo era en la no rotada, esto es, el 70,9. Aun así aparecen dos diferencias. Primero, la varianza se ha redistribuido para que el patrón de carga factorial y el porcentaje de varianza para cada uno de los factores sean distintos. Concretamente, en la solución rotada VARIMAX, el primer factor responde del 39,5 por ciento de la varianza en comparación con el 41,9 por ciento en la solución no rotada. De la misma forma, el segundo factor da cuenta de 31,4 por ciento frente al 29,0 por ciento en la solución no rotada. Así, el poder explicativo se ha desplazado ligeramente hacia una distribución más uniforme debido a la rotación. En segundo lugar, la interpretación de la matriz factorial se ha simplificado.

TABLA 3.7. Matriz de factores del análisis de componentes VARIMAX

	Cargas rotadas VARIMAX		
Variables	*Factor* 1	*Factor* 2	*Comunalidades*
X_1 Velocidad de entrega	–0,787	0,194	0,66
X_2 Nivel de precio	0,714	0,266	0,58
X_3 Flexibilidad de precio	–0,804	–0,011	0,65
X_4 Imagen del fabricante	0,102	0,933	0,88
X_6 Imagen del personal de ventas	0,025	0,934	0,87
X_7 Calidad de producto	0,764	0,179	0,62
			Total
Suma de cuadrados (autovalor)	2,37	1,88	4,25
Porcentaje de traza*	39,5	31,4	70,9

* Traza = 6,0 (suma de los autovalores).

Recuérdese que en la solución factorial no rotada todas las variables cargaban significativamente sobre el primer factor. En la solución factorial rotada, sin embargo, las variables X_1, X_2, X_3 y X_7 cargan significativamente sobre el factor 1, y las variables X_4 y X_6 cargan significativamente sobre el factor 2. Ninguna variable carga significativamente sobre más de un factor. Debería resultar aparente que la interpretación factorial se ha simplificado considerablemente al rotar la matriz factorial.

Denominación de los factores Cuando se ha obtenido una solución factorial satisfactoria, el investigador normalmente intenta atribuirle algún significado. El proceso implica la interpretación sustantiva del patrón de cargas factoriales para las variables, incluidos sus signos, en un esfuerzo por denominar cada uno de los factores. Antes de la interpretación, debe seleccionarse un nivel mínimo aceptable de significación para el factor. Normalmente, en la interpretación se emplean todas las cargas factoriales significativas. Pero las variables con mayores cargas influyen en una mayor medida en el nombre o etiqueta seleccionada para representar a un factor.

Observemos los resultados de la Tabla 3.7 para ilustrar este procedimiento. La solución factorial se obtuvo de un análisis de componentes con una rotación VARIMAX de las seis percepciones ofrecidas por HATCO. El punto de corte de las cargas a efectos interpretativos en este ejemplo fue del 55 por ciento (véase Tabla 3.2). Este resulta un punto de corte prudencialmente alto y puede ajustarse de ser necesario. Pero en nuestro ejemplo, todas las cargas caen sustancialmente por encima o por debajo de este umbral, haciendo la interpretación bastante fácil.

La interpretación sustantiva se basa en las mayores cargas significativas. El factor 1 tiene cuatro cargas significativas y el factor 2, dos. Para el factor 1, vemos dos grupos de variables. El primero es el nivel de precios (X_2) y la calidad del producto (X_7), ambas con signos positivos. Las otras dos variables, la rapidez en el envío (X_1) y la flexibilidad de precios (X_3) tienen signos negativos. Por ello, la calidad del producto y el nivel de precios varían juntos, lo mismo que la rapidez en el envío y la flexibilidad de precios. Sin embargo, un grupo respecto a otro se mueven en sentidos opuestos. En nuestro ejemplo, esto podría indicar que cuando calidad de producto y precio crecen, la rapidez en el envío o la flexibilidad de precios decrecen, o viceversa. Existen cuatro características tangibles de HATCO en el conjunto de variables, y están agrupadas juntas en un único factor. Este factor, quizá denominado *valor básico*, representa una concesión entre las percepciones del precio o calidad del producto y las percepciones de rapidez en el envío y flexibilidad de precios. Volviendo al factor 2, notamos que las variables X_4 (imagen del productor) y X_6 (imagen de los vendedores) se relacionan ambas con componentes de imagen, apuntando quizá a una etiqueta como *imagen HATCO* como segundo factor. Ambas variables tienen el mismo signo, sugirien-

do que estas percepciones son bastante similares en las respuestas no actuando en sentidos opuestos, como observamos con el primer factor.

Deberíamos notar que el servicio global (X_5) no estaba incluido en el análisis factorial. Cuando se presentaron las interpretaciones de las cargas factoriales, se decidió no incluir esta variable. Si se utilizan los resultados en otro análisis multivariante, X_5 podría incluirse como una variable aislada, aunque esto no aseguraría que fuera ortogonal a la puntuación factorial.

Ha quedado así expuesto el proceso de denominar los factores. Se habrá notado que está basado en la opinión subjetiva del investigador. En multitud de ocasiones, varios investigadores no dudarán en atribuir nombres diferentes a los mismos resultados debido a las diferencias en su formación y ejercicio. Por este motivo, el proceso de denominar los factores está sujeto a una controversia considerable. Pero si se puede asignar un nombre lógico que represente la naturaleza subyacente de los factores, normalmente facilita la presentación y comprensión de la solución factorial y por tanto es un procedimiento justificable.

La aplicación de una rotación oblicua La rotación VARIMAX es ortogonal, lo que significa que los factores permanecen incorrelacionados con el proceso de rotación. Pero en muchas ocasiones, los factores no necesitan estar incorrelacionados e incluso pueden estar conceptualmente ligados, lo que requiere una correlación entre los factores. En nuestro ejemplo, es bantante razonable esperar que las dimensiones perceptivas puedan estar correlacionadas; así quedaría justificada una rotación oblicua. La Tabla 3.8 contiene las matrices de patrón y estructura con las cargas factoriales para cada variable sobre cada factor. Como se discutió más arriba, la matriz de patrones se emplea normalmente con fines interpretativos, concretamente si los factores cuentan con una alta correlación entre ellos. En este caso, la correlación entre los factores es sólo del 0,12, por eso las matrices de patrones y estructura tienen cargas bastante comparables. Examinando las variables que cargan alto sobre cada factor, notamos que la interpretación es exactamente la misma que se encontró con la rotación VARIMAX.

TABLA 3.8. Rotación oblicua de una matriz factorial del análisis de componente

	Cargas de rotación oblicuas		
Variables	*Factor* 1	*Factor* 2	*Comunalidades*[a]
Matriz de patrones (cargas)			
X_1 Velocidad de entrega	–0,803	0,248	0,66
X_2 Nivel de precio	0,704	0,219	0,58
X_3 Flexibilidad de precio	–0,808	0,043	0,65
X_4 Imagen del fabricante	0,051	0,931	0,88
X_6 Imagen del personal de ventas	–0,026	0,937	0,87
X_7 Calidad de producto	0,759	0,129	0,62
Matriz de estructura			
X_1 Velocidad de entrega	–0,773	0,151	
X_2 Nivel de precio	0,730	0,304	
X_3 Flexibilidad de precio	–0,802	–0,054	
X_4 Imagen del fabricante	0,164	0,938	
X_6 Imagen del personal de ventas	0,088	0,934	
X_7 Calidad de producto	0,774	0,220	
Matriz de correlación de factores			
	Factor 1	Factor 2	
Factor 1	1,00		
Factor 2	0,121	1,00	

[a] Los valores de las varianzas compartidas no son iguales a la suma de las cargas al cuadrado debido a la correlación de los factores.

Sexto paso: *Validación del análisis factorial*

La validación de los resultados de cualquier análisis factorial es esencial, especialmente cuando se intenta definir la estructura subyacente entre las variables. Óptimamente, deberíamos seguir siempre nuestro análisis factorial con alguna forma de análisis factorial confirmatorio, tal como el modelo de ecuaciones estructurales (véase Capítulo 11), pero esto a menudo no es factible. Debemos contemplar otros medios, como el análisis de división de la muestra o la aplicación a muestras enteramente nuevas.

En este ejemplo, dividimos la muestra en dos partes iguales de 50 respuestas y reestimamos los modelos factoriales para contrastar la comparabilidad. La Tabla 3.9 contiene las rotaciones VARIMAX para los dos modelos factoriales, junto a las comunalidades. Como puede verse, las dos rotaciones VARIMAX son bastante comparables en términos tanto de las cargas como de las comunalidades para las seis percepciones. Un suceso notable es el cambio de signo del factor 1 en la primera submuestra frente a la segunda. Las interpretaciones de la relación entre las variables (por ejemplo, la rapidez del envío crece, la percepción del nivel de precios decrece) no cambian porque están relacionadas con las cargas en cada factor.

Con estos resultados, podemos estar más seguros de que los resultados son estables dentro de nuestra muestra. De ser posible, nos gustaría llevar a cabo un trabajo posterior obteniendo respuestas adicionales y asegurando que los resultados se generalizan entre la población.

Séptimo paso: *Usos adicionales de los resultados del análisis factorial*

El investigador tiene la posibilidad de utilizar el análisis factorial no sólo como instrumento para resumir datos, como hemos visto en la discusión previa, sino también como instrumento para reducir datos. En este contexto, el análisis de factores podría servir de ayuda para reducir el número de variables, o para seleccionar un conjunto de variables suplentes, una por factor, o creando nuevas variables compuestas para cada factor. Las siguientes secciones detallan los aspectos propios de la reducción de datos para este ejemplo.

Selección de variables suplentes para un análisis subsiguiente Examinemos los datos de la Tabla 3.7 para aclarar el procedimiento para la selección de variables suplentes. Prime-

TABLA 3.9. Validación del análisis de componente por estimación de división de la muestra VARIMAX

	Cargas rotadas VARIMAX		
Variables	*Factor* 1	*Factor* 2	*Comunalidades*
Muestras de división 1			
X_1 Velocidad de entrega	–0,695	0,397	0,64
X_2 Nivel de precio	0,772	0,142	0,62
X_3 Flexibilidad de precio	–0,822	–0,098	0,69
X_4 Imagen del fabricante	0,045	0,944	0,89
X_6 Imagen del personal de ventas	0,056	0,916	0,84
X_7 Calidad de producto	0,811	0,043	0,66
Muestras de división 2			
X_1 Velocidad de entrega	0,842	–0,002	0,71
X_2 Nivel de precio	–0,625	0,396	0,55
X_3 Flexibilidad de precio	0,829	0,107	0,70
X_4 Imagen del fabricante	–0,167	0,915	0,87
X_6 Imagen del personal de ventas	0,008	0,945	0,89
X_7 Calidad de producto	–0,681	0,315	0,56

ro, recordemos que las variables suplentes deberían seleccionarse sólo cuando la rotación sea ortogonal, porque cuando estamos interesados en utilizar variables suplentes en análisis subsiguientes, es lo mejor, en la medida de lo posible, que las variables independientes estén incorrelacionadas entre sí. Por ello debe elegirse una solución ortogonal en lugar de una oblicua.

Suponiendo que queremos seleccionar una única variable para su uso posterior, antes que construir una escala aditiva o utilizar la puntuación factorial (véanse las siguientes secciones), examinaríamos la magnitud de las cargas factoriales. Centrándonos en las cargas factoriales del factor 2, vemos que la carga para la variable X_4 es 0,933 y para la variable X_6 es de 0,934. La selección de un representante es difícil en casos como éste porque los tamaños de las cargas son esencialmente idénticos. Sin embargo, si no tenemos pruebas previas que sugieran que la fiabilidad o validación para una de las variables es mejor que para la otra, podríamos seleccionar la X_6 como variable suplente, sabiendo que representa elementos de imagen en mayor grado. Dadas las altas cargas para ambas variables, la selección de una sola podría ser suficiente debido al alto grado de intercorrelación entre ellas (mostrado por las altísimas cargas sobre el mismo factor o componente). De la misma forma, las cargas para el factor 1 son de 0,714 para la variable X_2 y de 0,764 para X_7, con cargas negativas comparables para X_1 (–0.787) y X_3 (–0.804). Para ambos factores, ninguna variable «representa» el mejor componente; por lo que la puntuación factorial o escalas aditivas serían más apropiados.

Creación de escalas aditivas Una escala aditiva es un valor compuesto para un conjunto de variables calculado por un procedimiento tan sencillo como tomar la media de las variables en la escala. Esto es bastante parecido a los valores teóricos en otras técnicas multivariantes, excepto que las ponderaciones para cada variable se supone son iguales en el proceso de promediar. El análisis factorial ayuda en la construcción de escalas aditivas identificando la dimensionalidad de las variables, que puede entonces estar relacionada con la definición conceptual. En este ejemplo, la solución de dos factores sugiere que deberían construirse dos escalas aditivas. Los dos factores, tratados previamente, corresponden a dimensiones que pueden denominarse y relacionarse con conceptos con un contenido de validación adecuado. La dimensionalidad de cada escala está apoyada por la interpretación «clara» de cada factor, con mayores cargas factoriales de cada variable sólo sobre un factor. La fiabilidad de las escalas aditivas está mejor representada por el alfa de Cronbach, que en este caso es de 0,77 para la escala 1 y de 0,85 para la escala 2. Ambos valores exceden el nivel recomendado del 0,70. Aunque no hay disponible ningún contraste directo para evaluar la validación de la escala aditiva, se pueden hacer comparaciones con análisis hechos con las variables originales y la puntuación factorial. La Tabla 3.10 ilustra el uso de las escalas aditivas junto con la puntuación factorial como sustitutos de las variables originales. Seleccionamos el ejemplo de identificar diferencias entre respuestas de pequeñas empresas frente a grandes (X_8). Las escalas aditivas muestran los mismos patrones de diferencias entre pequeñas y grandes empresas tanto en variables individuales como en la puntuación factorial. Por eso, demuestran cierto nivel de validación convergente con estas otras medidas.

Los diferentes signos de las cargas del factor 1 apuntan a un aspecto importante al construir escalas aditivas. Cuando las variables tienen cargas tanto positivas como negativas dentro del mismo factor, o bien las variables con las cargas positivas o negativas, deben tener los valores de los datos cambiados. Normalmente, las variables con las cargas negativas tienen la puntuación factorial cambiada por lo que las correlaciones, y las cargas, están ahora en positivo dentro del factor. La **puntuación inversa** es el proceso por el cual se cambian los valores de los datos de una variable para que las correlaciones con las otras variables cambien de signo (esto es, pasen de negativo a positivo). Por ejemplo, sobre nuestra escala de 0 a 10, podríamos cambiar la puntuación de una variable sustrayendo de 10 el valor original (es decir, puntuación inversa = 10 – valor original). De esta forma, las puntuaciones originales de 10 y 0 ahora tendrían las puntuaciones inversas de 0 y 10. Se mantendrían todas las características de la distribución; sólo que la distribución sería la opuesta.

La finalidad de la puntuación inversa es evitar una «cancelación» de variables con cargas positivas y negativas. Usemos como ejemplo dos variables con una correlación negativa, V_1 y V_2, con

TABLA 3.10. Valoración del reemplazo de las variables originales por puntuaciones de factores o escalas aditivas

Prueba estadística	*Diferencia de media entre grupos de encuestados basada en X_8, tamaño de empresa*			
	Puntuaciones de media		*Prueba-F*	
Medida	*Grupo 1: pequeñas empresas*	*Grupo 2: grandes empresas*	*Ratio F*	*Significación*
Variables originales				
X_1 Velocidad de entrega[a]	4,19	2,50	64,7	0,000
X_2 Nivel de precio	1,95	2,99	22,0	0,000
X_3 Flexibilidad de precio[a]	8,62	6,80	70,2	0,000
X_4 Imagen del fabricante	5,21	5,30	0,1	0,709
X_6 Imagen del personal de ventas	2,69	2,63	0,2	0,674
X_7 Calidad de producto	6,09	8,29	86,2	0,000
Puntuaciones de factores				
Puntuación de factor 1	–0,640	0,959	159,8	0,000
Puntuación de factor 2	0,052	–0,078	0,41	0,525
Escalas aditivas				
Escala 1	3,81	5,49	156,8	0,000
Escala 2	3,95	3,96	0,00	0,957

Correlaciones entre las puntuaciones de factores y las escalas aditivas

	Puntuaciones de factores		*Escalas aditivas*[b]	
	1	2	1	2
Puntuación de factor 1	1,000	0,000	0,995	0,075
Puntuación de factor 2	0,000	1,00	0,085	0,985
Escala aditiva 1[b]	0,995	0,085	1,000	0,154
Escala aditiva 2[b]	0,075	0,985	0,154	1,000

[a] Tienen cargas de factores negativas.
[b] Escalas aditivas calculadas como puntuación media a lo largo de los artículos. Por ejemplo, escala 1 es la media de X_1, X_2, X_3 y X_7.
Nótese: X_1 y X_3 son de escala inversa, debido a sus cargas de factores negativas.

V_1 teniendo carga positiva y V_2 negativa. Esto significa que si 10 el es la puntuación tope en V_1, el tope en V_2 sería 0. Supongamos ahora dos casos. En el caso 1, V_1 tiene un valor de 10 y V_2 de 0 (el mejor caso). En el caso 2, V_1 tiene un valor de 0 y V_2 de 10 (el peor caso). Si a V_2 no se le invierte la puntuación, entonces la escala de puntuación calculada sumando las dos variables para ambos casos 1 y 2 es 10, sin mostrar ninguna diferencia, mientras que sabemos que el caso 1 es el mejor y el caso 2 el peor. Si invertimos la puntuación de V_2, sin embargo la situación cambia. Ahora el caso 1 tiene valores de 10 y 10 sobre V_1 y V_2, respectivamente, y el caso 2 tiene valores de 0 y 0. La puntuación de escala aditiva es ahora de 20 para el caso 1 y 0 para el caso 2, que los distingue como la mejor y la peor situación.

Uso de las puntuaciones factoriales En lugar de calcular las escalas aditivas, podríamos calcular las puntuaciones factoriales para cada uno de los dos factores en nuestro análisis de componentes principales. De esta forma, cada respuesta contaría con dos nuevas variables (puntuaciones factoriales para los factores 1 y 2) que podrían sustituir a las seis variables originales en otras técnicas multivariantes. En el contraste de diferencia de medias entre dos grupos de respuestas (Tabla 3.9), vemos que todas las variables que cargan alto sobre el factor 1

(X_1, X_2, X_3 y X_7) son significativamente distintas en las respuestas de pequeñas y grandes empresas, mientras que las variables que cargan alto sobre el factor 2 (X_4 y X_6) no cuentan con diferencias significativas. Las puntuaciones factoriales y las escalas aditivas deberían mostrar patrones similares si son realmente representativas de las variables. Como se vió en la Tabla 3.10, las puntuaciones factoriales difieren en concordancia con este patrón. La puntuación factorial 1 muestra diferencias significativas que la puntuación del factor 2 no tiene. Se contemplan diferencias similares entre ambos grupos para las escalas aditivas. Además, estas escalas están altamente correlacionadas con las puntuaciones factoriales. Por eso, en este ejemplo, tanto las puntuaciones factoriales como las escalas aditivas retratan con precisión los conceptos que representan.

Selección del método de reducción de los datos Si las variables originales se van a reemplazar por las puntuaciones factoriales o las escalas aditivas, se debe decidir su uso. Esta decisión está basada en la necesidad de réplica en otros estudios (que favorece el uso de escalas aditivas) frente al deseo de ortogonalidad de las medidas (que favorece las puntuaciones factoriales). La Tabla 3.10 también contiene la matriz de correlación de las puntuaciones factoriales y las escalas aditivas. Debido a que empleamos una rotación ortogonal, la correlación entre las puntuaciones factoriales es de 0,0. Pero las escalas aditivas pueden estar correlacionadas y en este caso la correlación esa de 0,1545. El investigador debe averiguar la necesidad de la ortogonalidad frente a la replicabilidad al seleccionar las puntuaciones factoriales frente a las escalas aditivas.

Análisis factorial común: Pasos cuarto y quinto

El análisis factorial común es el segundo modelo de análisis factorial que presentamos. La primera distinción entre análisis de componentes principales y análisis factorial común es que el último considera sólo la varianza común asociada con un conjunto de variables. Este objetivo se cumple factorizando una matriz de correlación «reducida» con comunalidades iniciales estimadas en la diagonal en lugar de unidades. Las diferencias entre el análisis de componentes principales y el análisis factorial común aparecen en las etapas de estimación de los factores y de interpretación (pasos 4 y 5). Una vez que las comunalidades están en la diagonal, el modelo factorial común extrae los factores de un modo similar al análisis de componentes principales. El investigador utiliza los mismos criterios para la selección de los factores y su interpretación. Para ilustrar las diferencias que pueden aparecer entre análisis factorial común y de componentes principales, las siguientes secciones detallan la extracción e interpretación en el análisis factorial común de las seis percepciones HATCO empleadas en el análisis de componentes principales.

Cuarto paso: *Derivación de los factores y valoración del ajuste global*

Se ha empleado la matriz de correlaciones «reducidas» con las comunalidades en la diagonal en el análisis factorial común. Nótese que X_5 se omitió del análisis de componentes principales debido a un valor MSA inaceptable. Si se hubiera incluido en el análisis factorial común, sin embargo, la comunalidad no se podría haber estimado en la extracción original de factores. Por eso, el análisis factorial común todavía podría haberse llevado a cabo con seis variables incluso si X_5 no se hubiera eliminado debido al bajo valor MSA.

El primer paso es determinar el número de factores que se mantendrán para su examen y posible rotación. La Tabla 3.11 muestra las estadísticas de la extracción. Si fuéramos a emplear el criterio de raíz latente con un valor de corte de 1,0 para el autovalor, se mantendrían dos factores. Sin embargo, el análisis de caída indica que se retienen tres factores (véase Figura 3.10). Al combinar estos dos criterios, mantendremos dos factores para el análisis posterior debido al bajo

TABLA 3.11. Resultados para la extracción de factores comunes

Factor	*Autovalor*	*Porcentaje de varianza*	*Porcentaje acumulado de varianza*
1	2,51349	41,9	41,9
2	1,73952	29,0	70,9
3	0,59749	10,0	80,8
4	0,52956	8,8	89,7
5	0,41573	6,9	96,6
6	0,20422	3,4	100,0

valor del autovalor para el tercer factor y para mantener la comparabilidad con el análisis de componentes principales. De nuevo, como en el análisis de componentes principales examinado previamente, el investigador debería emplear una combinación de criterios para determinar el número de factores a mantener y podría incluso examinar la solución de tres factores como una alternativa.

La matriz de factores no rotados (Tabla 3.12) muestra que la solución factorial común da cuenta del 58,6 por ciento del total de la varianza. Debido a que el modelo factorial común final difiere a veces de las estimaciones de extracción iniciales (véase Tabla 3.11), el investigador debería asegurarse de evaluar las estadísticas de extracción para el modelo factorial común final. Si el investigador no estuviera satisfecho con la varianza total explicada, podría estimarse además un modelo de factores comunes que extraiga tres factores. Notemos que las comunalidades de cada variable son menores a las encontradas en el análisis de componentes principales. Esto se debe en primer lugar a la menor varianza global explicada, no al comportamiento de ninguna variable. De nuevo, la exploración del modelo de tres factores podría ser realizada para incrementar las comunalidades, junto con la varianza general explicada. Para nuestros propósitos aquí, interpretamos la solución de dos factores.

Quinto paso: *Interpretación de los factores*

Mediante el examen de las cargas no rotadas, señalamos la necesidad de una rotación de matriz factorial. Nos centramos entonces en el análisis factorial común rotado VARIMAX (Tabla 3.13). Examinemos cómo se compara con la matriz del análisis de componentes de factores rotados.

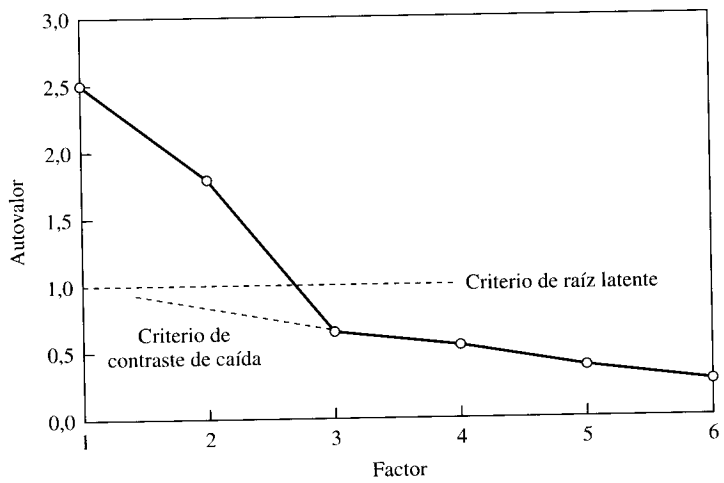


FIGURA 3.10. Contraste de caída para la solución del análisis factorial común.

TABLA 3.12. Matriz de factor común sin rotación

	Factores		
Variables	1	2	*Comunalidades*
X_1 Velocidad de entrega	0,485	0,512	0,50
X_2 Nivel de precio	0,629	0,187	0,43
X_3 Flexibilidad de precio	0,602	0,401	0,52
X_4 Imagen del fabricante	0,625	0,683	0,87
X_6 Imagen del personal de ventas	0,526	0,670	0,72
X_7 Calidad de producto	0,641	0,269	0,48
			Total
Suma de cuadrados (autovalor)	2,07	1,45	3,52
Porcentaje de trazo*	34,5	24,1	58,6

* Traza = 6,0 (suma de los autovalores).

La información que proporciona la solución de factores comunes es similar a aquella que proporciona la solución de análisis de componentes principales. Se proporcionan las sumas de cuadrados, los porcentajes de varianza, las comunalidades, las sumas totales de cuadrados y las varianzas totales extraídas de la misma manera que con la solución del análisis de componentes principales.

Una comparación de la información proporcionada en la matriz de análisis factorial común de factores y la matriz del análisis de componente de factores rotados demuestra una similitud notable. Las diferencias principales entre el análisis de componentes principales y el análisis factorial común son las cargas generalmente más bajas en el análisis factorial común, principalmente debido a las varianzas compartidas más bajas de las variables utilizadas en el análisis factorial común. Otra comparación que podría ser de utilidad para el investigador es el porcentaje de la varianza total indicado por cada factor. En el análisis de componentes principales (Tabla 3.7), los dos factores con rotación tenían una diferencia de un 8 por ciento (39,5 por ciento frente a 31,4 por ciento, respectivamente). En los resultados de factor común (Tabla 3.13), la de la rotación «dispersa» para los dos factores es casi igual en la varianza indicada (31,0 por ciento para el factor 1 y 27,6 por ciento para el factor 2). Sin embargo, incluso con estas diferencias tan ligeras en la varianza indicada, las pautas de las cargas y las interpretaciones básicas son idénticas entre el análisis de componentes principales y el análisis factorial común.

TABLA 3.13. VARIMAX matriz de factor común con rotación

	Cargas con rotación VARIMAX		
Variables	*Factor* 1	*Factor* 2	*Comunalidades*
X_1 Velocidad de entrega	–0,693	0,133	0,50
X_2 Nivel de precio	0,620	0,215	0,43
X_3 Flexibilidad de precio	–0,722	–0,026	0,52
X_4 Imagen del fabricante	0,109	0,925	0,87
X_6 Imagen del personal de ventas	0,037	0,846	0,72
X_7 Calidad de producto	0,677	0,155	0,48
			Total
Suma de cuadrados (autovalor)	1,86	1,66	3,52
Porcentaje de trazo*	31,0	27,6	58,6

* Traza = 6,0 (suma de los autovalores).

Una consideración gerencial de los resultados

Tanto el análisis de componentes principales como el análisis factorial común proporcionan al investigador varias ideas claves acerca de la estructura de las variables y las opciones para la reducción de datos. En primer lugar, con relación a la estructura de las variables, existen claramente dos dimensiones de valoración separadas y destacadas utilizadas por los clientes de HATCO. Una dimensión, denominada el *valor básico,* se refiere a los aspectos tangibles del HATCO y sus productos. Dentro de esta dimensión existe una concesión entre el precio del producto y la calidad frente a las características de la velocidad de entrega y la flexibilidad de precios. La segunda dimensión, la *imagen de HATCO,* se refiere a las percepciones de imagen del fabricante y su personal de ventas.

Los planificadores empresariales dentro de HATCO pueden debatir planes a partir de estas dos áreas en vez de tener que tratar con las variables por separado.

El análisis factorial también proporciona la base para la reducción de datos mediante las escalas aditivas o las puntuaciones de factores. El investigador tiene ahora un método para combinar las variables dentro de cada factor y con una puntuación única que puede reemplazar la serie de variables original con dos variables compuestas nuevas. Cuando se buscan diferencias, tales como aquellas que existen entre las empresas grandes y pequeñas, estas nuevas variables compuestas pueden ser utilizadas para que sólo se analicen las diferencias en dos valores, *el valor básico* e *imagen.*

Resumen

En este capítulo se ha presentado la técnica estadística multivariante del análisis factorial en términos conceptuales amplios. Se incluyen directrices básicas para interpretar los resultados para clarificar aún más los conceptos metodológicos. Se presenta un ejemplo de la aplicación del análisis factorial basada en la base de datos en el Capítulo 1.

El análisis factorial puede ser una técnica estadística multivariante muy útil y poderosa para la extracción efectiva de información de las grandes bases de datos y descifrar grandes cantidades de datos interrelacionados. Cuando funciona bien, indica relaciones interesantes que podrían no haber sido obvias mediante la simple investigación de los datos primarios o incluso la utilización de una matriz de correlación. El análisis factorial tiene la capacidad de identificar series de variables relacionadas e incluso desarrollar una única medida compuesta para representar la serie entera de variables relacionadas, con lo que se ofrece al investigador una herramienta poderosa para obtener una mejor comprensión de la estructura de los datos y una manera de simplificar otros análisis de una gran serie de variables mediante las variables compuestas de reemplazo. Las aplicaciones potenciales de las técnicas analíticas de factores para la solución de problemas y para la toma de decisiones en la investigación empresarial son numerosas. El uso de estas técnicas continuará creciendo conforme los investigadores adquieran familiaridad con los beneficios del resumen de datos y la reducción de datos.

El análisis factorial es una cuestión mucho más compleja y enrevesada de lo que podría parecer tras esta breve exposición. Tres de las limitaciones más citadas son las siguientes: primero, existen muchas técnicas para llevar a cabo los análisis factoriales y existe cierta controversia acerca de cuál de las técnicas es la mejor. En segundo lugar, los aspectos subjetivos del análisis factorial (por ejemplo, decidir cuántos factores extraer, qué técnica debería ser utilizada para rotar los ejes factoriales, cuáles de las cargas de factores son importantes) están todos sujetos a muchas diferencias de opinión. Tercero, el problema de fiabilidad es real. Igual que con otros procedimientos estadísticos, un análisis factorial empieza con una serie de datos imperfectos. Cuando los datos cambian como resultado de cambios en la muestra, el proceso de recogida de datos, o numerosos tipos errores de medida, los resultados del análisis también cambian. Los resultados de cualquier simple análisis son, por tanto, algo menos que perfectamente seguros. Este problema es especialmente

crítico dado que los resultados de una solución análitica de un único factor parecen frecuentemente factibles. Es importante destacar que la credibilidad no es garantía de validez o incluso estabilidad.

Cuestiones de revisión

1. ¿Cuáles son las diferencias entre los objetivos del resumen de datos y la reducción de datos?
2. ¿Hasta qué punto puede el análisis factorial ayudar al investigador a mejorar los resultados de las otras técnicas multivariantes?
3. ¿Cuáles son las directrices que se pueden usar para determinar el número de factores a extraer? Explíquelas brevemente.
4. ¿Cómo se usa la matriz de carga de factores para interpretar el significado de los factores?
5. ¿Cómo y cuándo se deben usar las puntuaciones de factores conjuntamente con las otras técnicas estadísticas multivariantes?
6. ¿Cuáles son las diferencias entre las puntuaciones de factores y las escalas aditivas? ¿Cuándo resulta más apropiada la aplicación de cada una de ellas?
7. ¿Cuál es la diferencia entre el análisis factorial tipo Q y el análisis cluster?
8. ¿Cuándo usaría el investigador una rotación oblicua en vez de una rotación ortogonal? ¿Cuáles son las diferencias entre ellas?

Bibliografía

1. Anderson, J. C., D. W. Gerbing, and J. E. Hunter (1987), «On the Assessment of Unidimensional Measurement: Internal and External Consistency and Overall Consistency Criteria.» *Journal of Marketing Research* 24 (November): 432–37.
2. American Psychological Association (1985), *Standards for Educational and Psychological Tests.* Washington, D.C.: APA.
3. Bearden, W. O., R. G. Netemeyer, and M. Moble (1993), *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior.* Newbury Park, Calif.: Sage.
4. Bentler, Peter M. (1992), *EQS Structural Equations Program Manual.* Los Angeles: BMDP Statistical Software.
5. BMDP Statistical Software, Inc. (1992), *BMDP Statistical Software Manual, Release 7*, vols. 1 and 2, Los Angeles: BMDP Statistical Software.
6. Borgatta, E. F., K. Kercher, and D. E. Stull (1986), «A Cautionary Note on the Use of Principal Components Analysis.» *Sociological Methods and Research* 15: 160–68.
7. Bruner, G. C., and P. J. Hensel (1993), *Marketing Scales Handbook, A Compilation of Multi-Item Measures.* Chicago: American Marketing Association.
8. Campbell, D. T., and D. W. Fiske (1959), «Convergent and Discriminant Validity by the Multitrait-Multimethod Matrix.» *Psychological Bulletin* 56 (March): 81–105.
9. Cattell, R. B. (1966), «The Scree Test for the Number of Factors.» *Multivariate Behavioral Research* 1 (April): 245–76.
10. Cattell, R. B., K. R. Balcar, J. L. Horn, and J. R. Nesselroade (1969), «Factor Matching Procedures: An Improvement of the s index; with tables.» *Educational and Psychological Measurement* 29: 781–92.
11. Chatterjee, S., L. Jamieson, and F. Wiseman (1991), «Identifying Most Influential Observations in Factor Analysis.» *Marketing Science* 10 (Spring): 145–60.
12. Churchill, G. A. (1979), «A Paradigm for Developing Better Measures of Marketing Constructs.» *Journal of Marketing Research* 16 (February): 64–73.
13. Cliff, N., and C. D. Hamburger (1967), «The Study of Sampling Errors in Factor Analysis by Means of Artificial Experiments.» *Psychological Bulletin* 68: 430–45.
14. Cronbach, L. J. (1951), «Coefficient Alpha and the Internal Structure of Tests.» *Psychometrika* 31: 93–96.
15. Dillon, W. R. and M. Goldstein (1984), *Multivariate Analysis: Methods and Applications.* New York: Wiley.
16. Dillon, W. R., N. Mulani, and D. G. Frederick (1989), «On the Use of Component Scores in the Presence of Group Structure.» *Journal of Consumer Research* 16: 106–12.
17. Gorsuch, R. L. (1983), *Factor Analysis.* Hillsdale, N.J.: Lawrence Erlbaum Associates.
18. Gorsuch, R. L. (1990), «Common Factor Analysis versus

Component Analysis: Some Well and Little Known Facts.» *Multivariate Behavioral Research* 25: 33–39.
19. Hattie, J. (1985), «Methodology Review: Assessing Unidimensionality of Tests and Items.» *Applied Psychological Measurement* 9: 139–64.
20. Joreskog, K. G., and D. Sorbo (1993), *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Mooresville, Ind.: Scientific Software International.
21. Kaiser, H. F. (1970), «A Second-Generation Little Jiffy.» *Psychometrika* 35: 401–15.
22. Kaiser, H. F. (1974), «Little Jiffy, Mark IV.» *Educational and Psychology Measurement* 34: 111–17.
23. McDonald, R. P. (1981), «The Dimensionality of Tests and Items.» *British Journal of Mathematical and Social Psychology* 34: 100–117.
24. Mulaik, S. A. (1990), «Blurring the Distinction Between Component Analysis and Common Factor Analysis.» *Multivariate Behavioral Research* 25: 53–59.
25. Mulaik, S. A., and R. P. McDonald (1978), «The Effect of Additional Variables on Factor Indeterminancy in Models with a Single Common Factor.» *Psychometrika* 43: 177–92.
26. Nunnally, J. L. (1978), *Psychometric Theory*, 2d ed. New York: McGraw-Hill.
27. Nunnally, J. (1979), *Psychometric Theory.* New York: McGraw-Hill.
28. Peter, J. P. (1979), «Reliability: A Review of Psychometric Basics and Recent Marketing Practices.» *Journal of Marketing Research* 16 (February): 6–17.
29. Peter, J. P. (1981), «Construct Validity: A Review of Basic Issues and Marketing Practices.» *Journal of Marketing Research* 18 (May): 133–45.
30. Robinson, J. P., P. R. Shaver, and L. S. Wrightsman (1991), «Criteria for Scale Selection and Evaluation,» in *Measures of Personality and Social Psychological Attitudes*, J. P. Robinson, P. R. Shanver, and L. S. Wrightsman (eds.). San Diego, Calif.: Academic Press.
31. Robinson, J. P., and P. R. Shaver (1973), *Measures of Psychological Attitudes.* Ann Arbor, MI: Survey Research Center Institute for Social Research, University of Michigan.
32. Rummel, R. J. (1970), *Applied Factor Analysis.* Evanston, Ill.: Northwestern University Press.
33. SAS Institute, Inc. (1990), *SAS User's Guide: Statistics, Version 6*. Cary, N.C.: SAS Institute.
34. Smith, Scott M. (1989), *PC-MDS: A Multidimensional Statistics Package*. Provo, Utah: Brigham Young University Press.
35. Snook, S. C., and R. L. Gorsuch (1989), «Principal Component Analysis versus Common Factor Analysis: A Monte Carlo Study.» *Psychological Bulletin* 106: 148–54.
36. SPSS, Inc. (1990), *SPSS Advanced Statistics Guide*. 4th ed. Chicago: SPSS.
37. Stewart, D. W. (1981), «The Application and Misapplication of Factor Analysis in Marketing Research.» *Journal of Marketing Research* 18 (February): 51–62.
38. Velicer, W. F., and D. N. Jackson (1990), «Component Analysis versus Common Factor Analysis: Some Issues in Selecting an Appropriate Procedure.» *Multivariate Behavioral Research* 25: 1–28.

LECTURA COMPLEMENTARIA

La imagen de la empresa detallista y su percepción por los consumidores: Una aplicación del análisis factorial

RODOLFO VÁZQUEZ CASIELLES
Universidad de Oviedo

ESIC-MARKET, n.º 65
Julio-agosto-septiembre 1989

I. INTRODUCCIÓN

La imagen juega dentro del Marketing un papel de gran importancia. A menudo se considera al concepto imagen como un determinante significativo de las ventas (e indirectamente de la rentabilidad) dado que permite al consumidor potencial adoptar una actitud favorable provocando una decisión o reacción positiva (de compra de un producto/marca, de mirar con buenos ojos a una empresa...).

Resulta razonable pensar que los establecimientos detallistas, al igual que la mayoría de las empresas, desean proyectar una imagen concreta mediante el desarrollo de su estrategia comercial. Así, observamos que cada tipo de minorista, presente en el mercado durante un periodo de tiempo, comienza a tener una «personalidad» o a transmitir significados a través de los que el consumidor los describe, recuerda o relaciona. La percepción por el consumidor de las empresas detallistas es lo que se conoce como imagen del detallista. Dicha imagen o «personalidad» es tan real como cualquier otro fenómeno psicológico y los consumidores deciden o no ir a comprar a los establecimientos detallistas en función de la imagen que tienen de los mismos.

En definitiva, una imagen de establecimientos detallistas podemos definirla como la impresión general de lo que una persona o grupo de personas piensa, percibe o sabe de estas fórmulas comerciales. La impresión general, en muchas ocasiones, es mas que una suma de atributos que caracterizan al negocio minorista. Sería necesario considerar que la interacci;on de todos estos atributos origina una percepción teóricamente distinta a la que se derivaría de cada atributo individualmente (efecto sinergia).

Por otro lado, los términos «posición o posicionamiento» han comenzado recientemente a utilizarse con frecuencia para significar imagen, excepto que implican un marco de referencia para la imagen fundamentándose en la competencia como punto de referencia: poner énfasis en la imagen definida en el contexto de interacción de varios atributos que caracterizan al detallista y tomando como referencia a los competidores que actúan en el mismo mercado.

Los detallistas tienen poder para configurar su imagen y posicionamiento en el mercado, dado que controlan gran parte de los estímulos físicos y psicológicos que influyen sobre el comportamiento del consumidor y determinan la naturaleza de la imagen minorista (véase Gráfico 3.L.1). Si la empresa detallista coordina cuidadosamente todos estos estímulos, probablemente conseguirá implantar en la mente de los consumidores la imagen deseada. Una imagen que debe ser atractiva, clara, incitante y que motive a los clientes potenciales hacia la compra en la propia empresa desechando la realizada en otros competidores.

Desde esta perspectiva en el presente estudio nos hemos propuesto:

1. Analizar el constructo imagen. Ello implica identificar los diversos factores qué integran la dimensión imagen y en qué medida influyen en crear una actitud favorable, atraer clientes, fomentar las compras y desarrollar juicios de satisfacción hacia los establecimientos detallistas.
2. Realizar una investigación empírica que permita, a partir de las conclusiones previas, medir la imagen de diferentes tipos de detallistas y derivar su posicionamiento competitivo en un mercado geográfico concreto.

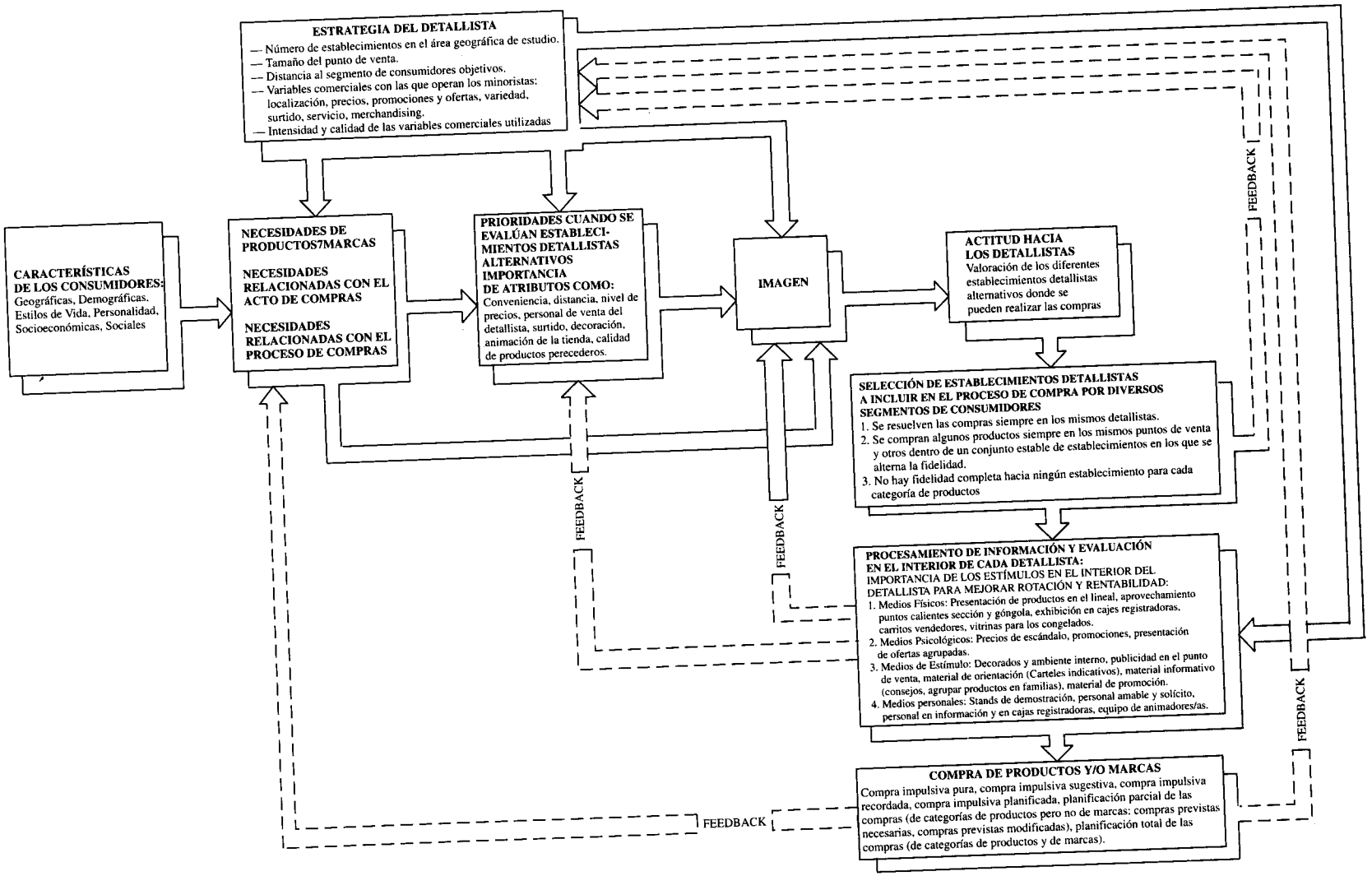


GRÁFICO 3.L.1. El proceso de selección y de compra en los detallistas por los consumidores.

II. DIMENSIONES RELEVANTES DE LA IMAGEN DE LOS DETALLISTAS

Distintos estudios examinan el conjunto de factores significativos considerados por el consumidor cuando evalúa a los detallistas (1). Sus objetivos consisten en identificar los componentes principales de la imagen, señalar sus interrelaciones y detectar su importancia en la selección de establecimientos (SPIGOLE y SEWALL, 1987), formación de preferencias y niveles de satisfacción (WESTBROOK, 1981). En algunos casos pretenden establecer las diferencias existentes en la evaluación de los componentes de la imagen en orden a desarrollar tipologías de consumidores (HAWKINS: BEST y CONEY, 1986). Incluso se admite que dichos componentes son idóneos para diversas áreas geográficas, aunque su importancia varía por tipo de detallistas y categorías de productos vendidos (HIRSCHMAN: GREENBERG y ROBERSTON, 1978). Globalmente, todos los estudios afirman que los consumidores elaboran juicios y seleccionan establecimientos detallistas mediante puntuaciones subjetivas a los componentes de la imagen.

En esta sección comentamos brevemente las dimensiones que integran la imagen de los detallistas (de forma individual y conjunta) indicando en qué medida son interesantes para una estrategia comercial, puesto que permiten influir en las preferencias de los consumidores y alcanzar un determinado volumen de ventas respetando los objetivos de rentabilidad (véase Gráfico 3.L.2).

II.1. Características de los productos vendidos por el detallista

El surtido de productos ofertado por un detallista es, no cabe duda, quien en gran medida «hace la tienda, define su personalidad y asegura el éxito». Constituye un factor esencial de la vocación del establecimiento, de la imagen que éste ofrece y de las preferencias del consumidor. El surtido o selección de productos a vender es siempre un compromiso entre la amplitud (número de líneas o familias de productos comercializados) y la profundidad (cantidad de artículos o referencias en cada familia).

Son múltiples los factores que determinan el surtido (SALEN, 1987) y de cuya interacción resulta una imagen favorable o deseada (véase Gráfico 3.L.3). En particular mencionamos: vocación del detallista, demanda del mercado, mayor o menor afluencia de distintas categorías de clientes con necesidades y hábitos propios, el entorno comercial que crea competencia para ciertos artículos y brinda oportunidad para otros, la superficie de venta y el capital de que se dispone para inversiones fijas (mobiliario, instalaciones) y para el stock (importancia de la rotación).

El precio asociado con los artículos que componen el surtido (principalmente para los productos considerados como de elevada atracción), puede parecer en principio un determinante significativo de la imagen y de las preferencias hacia los establecimientos detallistas. Sin embargo, las investigaciones sobre la importancia del precio en la selección de detallistas ofrecen resultados contradictorios, puesto que varían con la clase de productos, el tipo de minorista y la clientela.

Por un lado, podemos afirmar que cuando las diferencias «percibidas» (2) en los precios de los mismos productos en distintos detallistas son apreciables, el precio se convierte en un atributo de decisión importante (DELLA BRITTA: MONROE y McGINNIS, 1981: DELLA BRITTA y MONROE, 1981). Por otro lado, también reconocemos que el precio está subordinado a otras variables. Así, un establecimiento detallista que desarrolla una elevada reputación sobre expansión y calidad de los productos con una buena política de «servicios» y un personal de ventas atento, amable y solícito, puede encontrar clientes que estén dispuestos a pagar una cantidad adicional por disfrutar de dichas ventajas cuando realizan sus compras. En definitiva, interesa utilizar el precio como arma competitiva siempre que la elasticidad de la demanda así lo aconseja y cuando la empresa sea capaz de controlar sus gastos para no poner en peligro la rentabilidad.

La promoción de ventas es una herramienta básica para atraer al consumidor hacia el establecimiento detallista y estimular las ventas de otros productos no promocionados. Ahora bien, fomentar la promoción como el atributo más relevante que configura la imagen de la empresa puede resultar un factor inadecuado si con prioridad no se analizan sus efectos sobre los beneficios totales del minorista.

Los diseños experimentales como el «Cuadro Latino» o «Factorial», pueden ser utilizados para captar la influencia de la promoción (y su interacción con otras variables de merchandising), aunque muchas veces tan sólo sobre las ventas de los productos promocionados (WILKINSON: MASON y PAKSOY, 1982). La técnica de los paneles de consumidores o de familias permite además, a partir de la acumulación de los datos ofrecidos por los panelistas, cuantificar el incremento o disminución de las ventas (o de los beneficios) debido a un cierto número de variables analizadas por distintos autores: sustitución de otras marcas por productos en promoción (WALTERS y RINNE, 1986; KUMAR y LEONE, 1988), mayores compras de mercancía complementaria a la promocionada (LEWISON y DELOZIER, 1986) o de

(1) Para una revisión de la literatura sobre el tema consultar MAZURSKY y JACOBY (1986); HIGIE, FEICK y PRICE (1987).

(2) El precio es un concepto relativo. El consumidor tan sólo considera una serie de establecimientos («conjunto evocado») como alternativas de elección cuando se plantea una compra. Dicho conjunto evocado varia con el tiempo y es función de la experiencia, del tipo de productos adquiridos, de las comunicaciones a las que está expuesto el consumidor y de su grado dé implicación con el acto de compra. En el mejor de los casos el consumidor compara precios entredetallistas pertenecientes al conjunto evocado. Además, los consumidores no conocen el precio exacto de la mayoría de los productos debido a que éstos presentan constantes variaciones y a que los individuos son incapaces de tratar de manera satisfactoria una «carga informativa» elevada. Tan sólo son conscientes de unos precios de referencia (en virtud de su experiencia y la intensidad publicitaria existente) admitiendo bandas de fluctuación en torno a los mismos antes de modificar sus hábitos de compra.

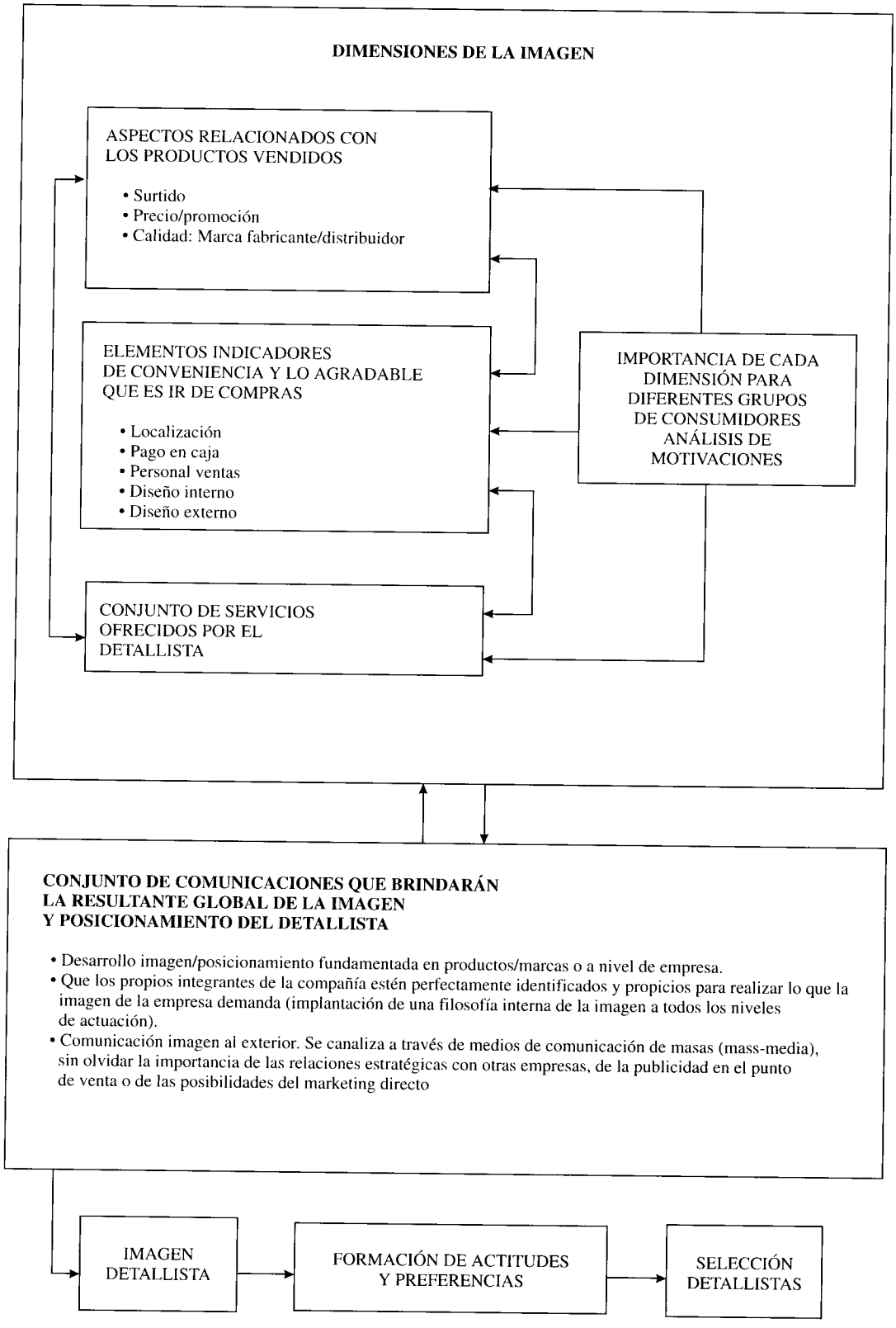


GRÁFICO 3.L.2. Factores determinantes imagen detallistas.

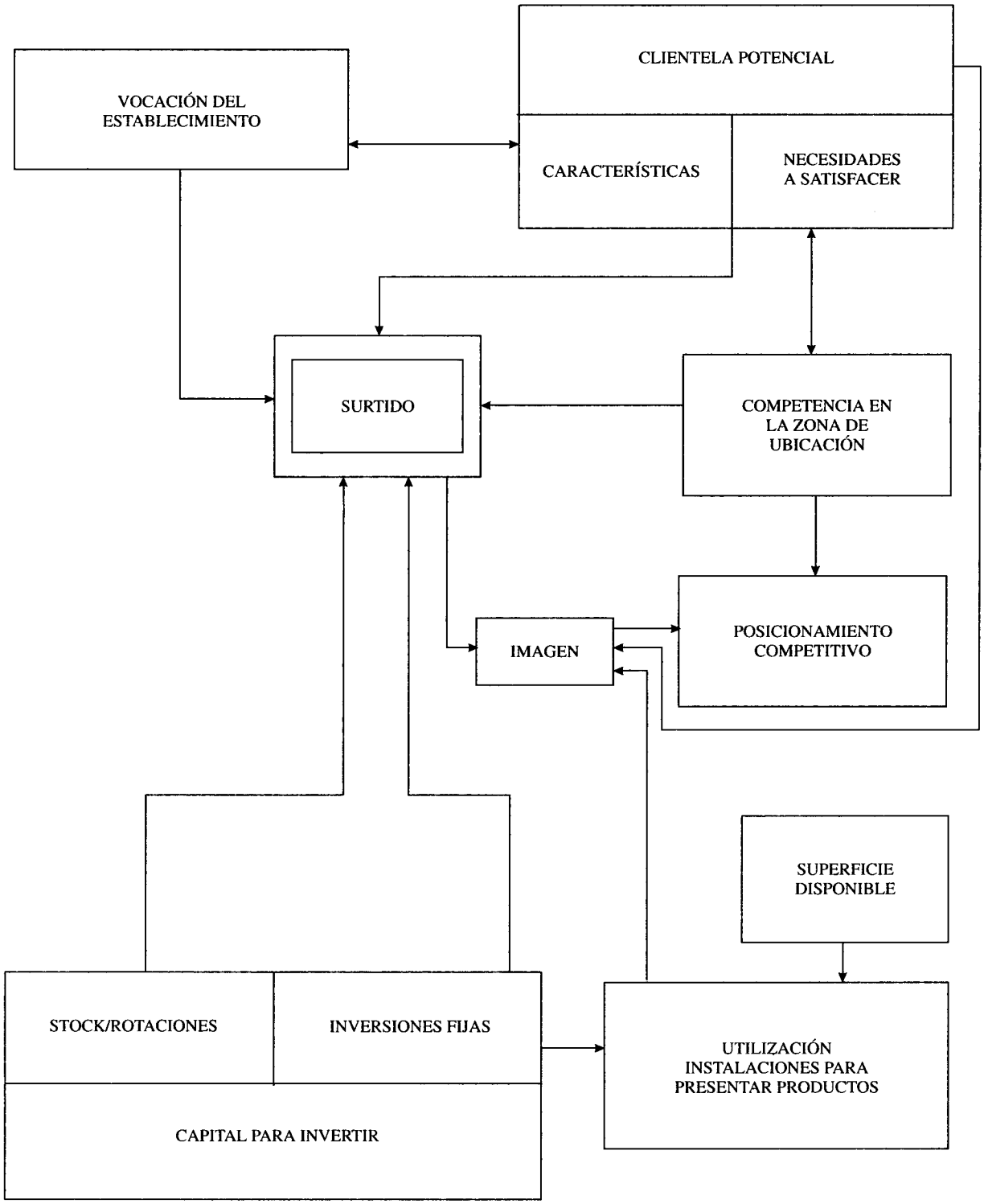


GRÁFICO 3.L.3. Factores que determinan el surtido y su influencia sobre la imagen.

artículos no relacionados con la misma para aprovechar adquisiciones múltiples en un solo desplazamiento de compra (OLSHAVSKY y CRASK, 1983) y en definitiva una más elevada atracción sobre los clientes habituales del establecimiento y sobre clientes que frecuentan regularmente otros detallistas (DUNCAN, HOLLANDER y SAVITF, 1983). Otras investigaciones más recientes estiman las interrelaciones comentadas mediante enfoques de modelización causal y una complicada metodología de ecuaciones estructurales cuya solución se obtiene aplicando programas como el LISREL-VI (WALTERS, 1988). La disponibilidad de datos obtenidos con un scanner facilita este tipo de estudios.

Las investigaciones propuestas no están exentas de problemas. Por ejemplo, se necesitan períodos de ausencia de promociones para contrastar resultados y esto implica importantes costes de oportunidad en relación a otros detallistas principalmente por no poder responder a la competencia. Si la empresa es capaz de solventarlos y detectar efectos positivos de la promoción sobre las ventas totales y el beneficio, estaría interesada en proyectar una imagen centrada en factores promocionales. Incluso, los beneficios pueden entenderse más allá del período promocional: aumenta la competitividad a largo plazo; se crean barreras desanimando la entrada de nuevos competidores para así mejorar la lealtad hacía el detallista e incrementar las compras repetidas en periodos posteriores a la compra.

Una parte esencial de la estrategia de los distribuidores es el desarrollo de marcas propias (3). Constituye una forma adicional de alcanzar reputación, crear lealtad del cliente hacia el establecimiento, construir una imagen corporativa y diferenciarse con respecto a los competidores.

El éxito de fomentar una imagen rentable mediante las marcas propias estriba en (DIJKSTRA TAUREL, 1987):

1. Adoptar una estrategia ofensiva orientada a ofrecer «valor» al consumidor (la calidad de lo que se compra, por el precio al que se compra), posicionando las marcas propias como una alternativa a las marcas nacionales líderes en cada línea de productos. Silos consumidores perciben las marcas propias como de aceptable calidad (contrastada en laboratorios por especialistas en tecnología de los alimentos y nutrición) estarán dispuestos a adquirir productos de precio más barato.
2. Abandonar una estrategia defensiva consistente en «seguir muy de cerca al líder» de las marcas nacionales, lanzando una «copia» de cualquier producto novedoso inmediatamente después de que lo hace el fabricante. Las marcas propias deben tomar la iniciativa investigando estilos de vida, apreciación del valor que mueve a los consumidores, posicionarse e identificarse en el mismo nivel que las marcas nacionales líderes, en vez de ofertar solamente como una alternativa más barata (una imagen de economía). Sólo así se conseguirá transmitir una imagen fuerte y familiaridad con los productos vendidos. Todo ello sin olvidar la adecuada presentación en envases atractivos.

(3) Empleamos la terminología tradicional de marcas propias o «marcas controladas por el distribuidor» y marcas nacionales o «marcas controladas por el fabricante», aunque en ocasiones las marcas propias tienen una mayor amplitud de distribución geográfica que las marcas nacionales que son de carácter local. Ello permite distinguir entre marcas nacionales lideres en cada línea de productos y otras pequeñas marcas nacionales o locales de menor volumen de ventas. También diferenciamos entre marcas propias y productos genéricos (HARRIS y TRANO, 1985: McENALLY y HAWES, 1984). A estos últimos no se les asigna un nombre de marca que avale su presentación, el envasado/empaquetado es más sencillo y eliminan costes de promoción publicidad, lo que permite trabajar con precios más baratos que las marcas propias.

II.2. Factores que influyen en la conveniencia y en lo agradable que puede ser ir de compras

Hasta aquí ha quedado patente que los factores decisivos de la compra que permiten asegurar la fidelidad de una clientela son: precio, calidad y surtido. Sin embargo, existe otro conjunto de variables que pueden ser utilizadas para fomentar la imagen de la empresa detallista dado que ayudan a lograr clientes satisfechos, aumentan el importe de las compras de una clientela adicta y aportan nuevos clientes debido a la reputación de la oferta realizada. Hacemos alusión a aspectos relacionados con la conveniencia/comodidad de ir de compras y con el diseño externo e interno del establecimiento minorista. Se trata de factores determinantes en la elección del lugar de compra por el consumidor siempre que (RUNYON y STEWART, 1987):

1. Hay un número elevado de competidores en la zona geográfica de estudio.
2. Existen relativamente pocas diferencias en precios, surtido y calidad de los productos vendidos.
3. Es factible presentar la oferta de diversas maneras alternativas puesto que existen consumidores con múltiples necesidades, pertenecientes a diferentes clases sociales y con distintos estilos de vida, que disfrutan con el valor de la estimulación y novedad desarrolladas por los establecimientos detallistas.

La conveniencia proporciona ahorro al consumidor en términos de inversión de tiempo en el proceso de compra. Indirectamente implica comodidad. Incluye factores relacionados con la localización del establecimiento detallista, el pago en caja, el horario y otras variables vinculadas al diseño del minorista (distribución por secciones y ubicación de productos en cada sección de forma que permita una compra rápida) y a la disponibilidad de productos (el coste de quedarse sin stock incluye la pérdida del producto y también el tiempo de su búsqueda).

Las ventajas de conveniencia asociadas a la localización de un detallista incorporan, además de la distancia física al establecimiento, la accesibilidad, disponibilidad de aparcamiento y características de la zona comercial circundante al minorista, donde también se concentran otras diversas ofer-

tas de servicios comerciales (4). El detallista distanciado del domicilio de los clientes debe desarrollar una estrategia de imagen que enfatice factores tendentes a reducir la distancia percibida, subrayar las facilidades de aparcamiento, facilitando el desplazamiento al lugar donde se localiza el establecimiento, indicando la posibilidad de realizar otras compras complementarias.

Una mayor conveniencia y comodidad también implica la necesidad de que la empresa sea ingeniosa para reducir la espera ante las cajas al mínimo, equipándose con un material idóneo de administración. El sistema de lectura óptica a través de scanners es un medio para acelerar el proceso de «compra-pago-salida». El check Robot ACM es un nuevo concepto que permite a los consumidores realizar el control de salida de sus compras (SALEN, 1988). En establecimientos americanos y franceses se presenta un solo ejemplar de cada articulo en venta. El cliente se provee de tantas tarjetas especiales (situadas en una caja, bajo cada producto) como artículos deseados y los entrega a una cajera para que ésta las introduzca en una tabuladora y un departamento especial se encargue de preparar y empaquetar el pedido mientras se efectúa el pago.

Los diseños interno y externo influyen en la imagen del detallista, en su habilidad para atraer y retener clientes, proporcionan el entorno en el que se deciden las restantes variables estratégicas y permiten alcanzar una ventaja competitiva, sobre todo cuando surtido y precio son similares.

El diseño externo hace referencia a las características de la fachada, con rótulos sobre razón social o nombre del establecimiento, a la manera de presentar los productos en los escaparates y a la forma de la entrada (SALEN, 1987). El diseño exterior ha de facilitar la identificación del establecimiento y el comerciante debe analizar alternativas de imagen dando respuesta a interrogantes como los siguientes: ¿es visible para el tráfico en automóvil o a pie? ¿Necesita ser distinto de otros establecimientos circundantes? ¿Ha de ser similar a otros locales de la misma empresa? ¿Desde el exterior se identifica la mercancía que hay dentro? ¿Da la impresión de estabilidad y permanencia? ¿El coste de construcción y mantenimiento es razonable? ¿Necesita fomentar una imagen moderna, progresiva o conservadora? ¿Los colores crean sentimientos de excitación y movimiento o de relajación y serenidad?

El entorno interno o «atmósfera del detallista» afecta al placer de ir de compras, a la inclinación de visitar el establecimiento detallista y pasar bastante tiempo en el mismo, particularmente para aquellos segmentos de consumidores con fuertes necesidades de estimulación y novedad que disfrutan con el acto recreativo de la compra, con la interacción social que se produce en el acto de compra o con poder adquirir experiencia en la compra de productos (BELLENGUER y KORGAONKAR, 1986).

Algunos elementos que definen el diseño interior son (ANDRES, 1983: MASSON y WELLHOFF, 1984): mobiliario (góndolas, anaqueles, mostradores, vitrinas) y equipamiento (cajas registradoras, ascensores, escaleras automáticas); iluminación para presentar la mercancía de manera efectiva; disposición de las secciones y ubicación de productos facilitando, mediante diferentes pasillos, la circulación de los clientes por la totalidad de los lineales; aprovechamiento de áreas muertas; stands de demostración; zonas para poder comprar con tranquilidad; ambientación sonora y decoración; peculiaridades del personal de ventas (en cuanto a su apariencia, cordialidad, cortesía, amabilidad y servicialidad); material de orientación (carteles indicativos, flechas) y de información (sobre familias de productos, su utilización, consejos).

En definitiva, los efectos de la imagen mediante un adecuado entorno interno, se dirigen a incentivar la compra y a que ésta sea agradable, pero tienen el inconveniente de que al ser de carácter emocional a menudo son difíciles de verbalizar por los clientes (DONOVAN y ROSSITER, 1982).

II.3. Conjunto de servicios ofrecido por el detallista

Conceptualizamos el servicio como una ampliación de la oferta del detallista y no como una noción genérica (similar a la expuesta en el trabajo de HUMMEL y SAVITT, 1988) que integra el sistema global de satisfacción al consumidor (resultado de todos los elementos que configuran la pretransacción, transacción y post-transacción entre la empresa y sus clientes).

Se trata de servicios que, al proporcionar también conveniencia y comodidad al consumidor, influyen en la imagen proyectada por el detallista. En particular mencionamos: servicios financieros (aceptación de tarjetas de crédito propias o bancarias); servicios complementarios en el momento de compra (embalaje de productos, regalo de bolsas); aceptación de pedidos por teléfono con acuse de recibo e indicación de un plazo de entrega; métodos de reparto de la mercancía vendida formando al personal para que esta acción produzca una impresión positiva; mantenimiento de contactos para informar de posibles ofertas y analizar reclamaciones por productos y clientes; implantar un sistema de cambios de productos defectuosos; servicios de instalación, puesta en funcionamiento y mantenimiento de productos comprados (a veces se requieren contratos con terceros para asegurar esta tarea). En conclusión, es necesaria la puesta en marcha de una organización especial, cuya rentabilidad ha de ser calculada, para que este tipo de servicios equilibre gastos o produzca ligeros beneficios.

---

(4) Cuando el consumidor opta por desplazarse fuera de su lugar de residencia (debido a la insuficiencia o al escaso atractivo de la oferta local) considera diversos aspectos:

a) Características de establecimientos ubicados en el centro comercial de destino y en la ruta que ha de seguir para llegar a dicha localización: en qué medida su surtido permite aprovechar el desplazamiento para realizar diferentes compras en distintos detallistas; facilidades de aparcamiento y para desplazarse de un establecimiento a otro en el mismo centro comercial; horario de apertura al público; ausencia de multitud de personas que dificulten la compra (BLACK y REILLY, 1983).

b) Factores relacionados con el transporte: tiempo de viaje; opciones disponibles de transporte, así como su coste, confort y seguridad (ROWES y LUMPKIN, 1984).

II.4. Interrelaciones entre los factores que integran la imagen

Los elementos determinantes de la imagen y selección de establecimientos detallistas implican una relación compleja entre todos los factores estudiados. Para clarificar dicha constelación de interrelaciones se han venido utilizando diferentes procedimientos metodológicos como los modelos de actitud multiatributo (JAMES: DURAND y DREVES, 1976), tratamiento de escalas multidimensionales (SINGSON, 1975), análisis de regresión (SCHIFFMAN: DASH y DILLON, 1977), análisis factorial (HIRSCHMAN; GREENBERG y ROBERSTON, 1978), análisis discrimante (PESSEMIER, 1980), análisis conjunto (SHARMA y MALHOTRA, 1981) y enfoques logit o probit (MAHOTRA, 1983). Todos ellos tienen como objetivo detectar las características más importantes que configtiran la imagen de las empresas minoristas. En este artículo proponemos escalas de actitud y análisis factorial para representar la imagen de diferentes tipos de detallistas.

II.5. Tipología del consumidor

Las características de los productos vendidos, los factores que influyen en lo agradable que es ir de compras y las políticas de servicios, aunque son de gran utilidad, no proporcionan referencias exactas sobre las motivaciones del consumidor para ir de compras ni sobre la forma en que estas motivaciones inciden sobre la imagen y selección de establecimientos detallistas. En este apartado pretendemos comprender el comportamiento de los consumidores en términos de sus actitudes hacia la compra. Examinamos cuatro tipologías del consumidor particularmente relevantes para conocer cómo diferentes segmentos de mercado pueden percibir imágenes de los detallistas y valorar (o desear) de manera distinta los elementos que la integran (HAWKINS: BEST y CONEY, 1986):

1. Tipología fundamentada en el riesgo percibido. La compra de los productos implica cierto riesgo que puede incluir tanto consecuencias económicas como sociales. Ciertos productos, a causa de su precio elevado o de su complejidad técnica, representan altos niveles de riesgo económico. Productos intrínsecamente relacionados con la imagen pública del consumidor presentan altos niveles de riesgo social. La percepción del riesgo difiere entre consumidores debido principalmente a sus experiencias pasadas y a su estilo de vida. Por esta razón, el riesgo percibido es considerado una característica del consumidor más bien que una característica del producto. Por ejemplo, una forma de reducir el riesgo social es comprar en establecimientos detallistas que connotan el prestigio social deseado por el consumidor.
2. Tipología fundamentada en la autoconfianza del consumidor. Los consumidores con más seguridad en que las decisiones que toman son las acertadas es más probable que acudan a comprar a nuevos detallistas. En caso contrario, se dirigirán a minoristas suficientemente conocidos quienes, al menos mentalmente, reducen el riesgo de compras insatisfactorias. La combinación de elevados niveles de riesgo y reducida confianza en uno mismo, experimentada muchas veces para bienes duraderos, sugiere que los detallistas deberían reducir incertidumbre, ofertando marcas conocidas, realizando campañas de promoción o enfatizando la asistencia que el personal de ventas ofrecerá sobre las características de los productos vendidos para facilitar la elección más adecuada.
3. Tipología relacionada con las características familiares. Es frecuente encontrarse con situaciones donde un miembro de la familia selecciona los productos o marcas y otro miembro diferente selecciona el establecimiento detallista donde realizar las compras. También puede darse una decisión conjunta cuando las compras se realizan por varios miembros de la familia que incluso acuden siempre unidos al minorista. La etapa en el ciclo de vida de las personas y si la mujer trabaja o no, presentan importantes influencias sobre el proceso de compra. Por ejemplo, las mujeres que trabajan compran con mayor frecuencia los fines de semana que las amas de casa, perciben de manera distinta los establecimientos detallistas y asignan más tareas de compra a otros miembros de la familia.
4. Tipología que toma como referencia la orientación a la compra. Los individuos van a comprar por razones mucho más complejas que la simple adquisición de productos (WESTBROOK y BLACK, 1985). Características como evadir las actividades rutinarias, estimulación sensorial, interrelaciones sociales, aprendizaje de nuevas tendencias e incluso adquisición de poder interpersonal, han sido mencionadas como razones no expresamente relacionadas con las necesidades de productos y que son tenidas en cuenta cuando se compra. Por supuesto, la importancia de estos motivos varía según los individuos considerados y para un mismo individuo a medida que transcurre el tiempo y la situación cambia (véase Gráfico 3.L.4).

III. INVESTIGACIÓN EMPÍRICA SOBRE IMAGEN Y POSICIONAMIENTO COMPETITIVO DE ESTABLECIMIENTOS DETALLISTAS (5)

Tomando como referencia las puntualizaciones conceptuales expuestas con prioridad, optamos por investigar alguno de los aspectos que configuran la imagen de los establecimientos detallistas de alimentación. La investigación se realizó

(5) Un resumen de esta sección fue presentado en el III Congreso de la Asociación Española de Economía de la Empresa (AEDEM) celebrado en Bilbao los días 22, 23 y 24 de mayo de 1989.

1. ECONOMÍA EN ESTABLECIMIENTO Compara precios, aprovecha promociones, regatea en la compra. Piensa que puede ahorrar repartiendo compras entre distintos detallistas. La compra deultramarinos no es tarea importante, ni una oportunidad para evadir la rutina. Desean detallistas con precios especiales y marcas del distribuidor con calidad. No muy implicados en la búsqueda o planificación previa a la compra.	**4. PLANIFICADOR ECONÓMICO** Actitud positiva hacia la comparación de precios y el aprovechamiento promociones. Fuertemente implicado en la planificación previa compra analizando publicidad sobre precios unitarios. Sus atributos ideales son la calidad de marcas del distribuidor y conveniencia. No les gusta tratar con nuevas marcas ni cambiar detallista.
2. APÁTICO/MECANICISTA Escaso entusiasmo por el acto de ir de compras. Perciben poca utilidad en las actividades de planificación previa a lacompra. Poco interés por reclamoscompra.	**5. IMPLICADO/TRADICIONAL** Disfruta con el acto de ir de compras, comparando precios y promociones. Fuertemente planificadores. La compra es un acto recreativo.
3. CONVENIENCIA No comparan precios, aunque aprovechan promociones. No les gusta visitar otros detallistas puesto que conocen la ubicación exacta de los productos en los establecimientos donde compran en la actualidad. Orientados hacia marcas líderes nacionales dado que para ellos implica calidad. Atributo más importante conveniencia. Escaso interés por marcas distribuidor. Reducida planificación previa a la compra.	**6. HOGAREÑO/FAMILIAR** Creen que el nombre de marca fabricante implica calidad. Piensan que la compra es una tarea importante, pero no disfrutan con actividades de planificación previa a la compra. No son innovadores, pues tienen actitudes negativas hacia la compra en más de un detallista. Menor interés por publicidad precios. Influencia amistades.

Fuente: Adaptado de GUILTINAN y MONROE (1980).

GRÁFICO 3.L.4. Tipología consumidores de acuerdo con su orientación a la compra.

durante los meses de enero y febrero de 1989, mediante el desarrollo y cumplimentación de un cuestionario. El estudio se circunscribe al ámbito geográfico de Asturias y más específicamente a las comarcas de su Zona Central.

La muestra utilizada es de 500 encuestas válidas a consumidores después que éstos realizaran sus compras en diversos tipos de establecimientos detallistas (un nivel de confianza del 95,5 por ciento y un error deseado del 4,47 por ciento). El método empleado para el reparto de la muestra fue el «estratificado con afijación proporcional» por comarcas. En cada comarca se repartieron las encuestas de tal forma que estuvieran representados los clientes, conforme a diversos intervalos de edad, de: todos los hipermercados; al menos un establecimiento de todas las cadenas comerciales de superservicios-supermercados; un economato; las encuestas restantes se asignaron a compradores de autoservicios y tiendas tradicionales (respetando la relación dos a tres).

El cuestionario integraba preguntas sobre frecuencia de compra, lugar donde se compra, razones para no acudir siempre a los mismos detallistas, motivos de compra (6), escalas de actitud para valorar establecimientos minoristas y determinar su posicionamiento competitivo y variables de clasificación de los consumidores. A continuación presentamos los resultados más relevantes en relación a la imagen de los detallistas (7).

III.1. Percepción de los establecimientos detallistas: Estimación de perfiles de imagen

Se procedió inicialmente a determinar la serie de atributos más relevantes por medio de los cuales se perciben y evalúan los diferentes detallistas. La revisión de investigaciones similares (comentadas con prioridad) diversas entrevistas en profundidad y una reunión de grupo, facilitaron la elaboración de la escala de actitud propuesta en el Cuadro 3.2.1 e incorporada al cuestionario. Dicha escala fue presentada a los consumidores solicitando valorasen las característica del minorista donde realizaban sus compras.

(6) En todos los casos la información se desglosó para diferentes categorías de productos: alimentación, bebidas, carnes, charcutería, frutas, verduras, droguería, limpieza y perfumería.

(7) Para un análisis completo de todo el estudio el lector interesado puede consultar VÁZQUEZ CASIELLES y TRESPALACIOS (1989).

CUADRO 3.L.1. Escala de actitud para el análisis de imagen y posicionamiento de establecimientos detallistas

Código	*Atributos*	*Grado de acuerdo*				
		5	*4*	*3*	*2*	*1*
V1	Es una tienda agradable para comprar.					
V2	El establecimiento realiza muy buenas promociones y ofertas.					
V3	El establecimiento está limpio y cuidado.					
V4	Resulta fácil encontrar el artículo deseado.					
V5	Los precios son más baratos que en otros establecimientos que venden productos similares.					
V6	El personal es amable y solícito.					
V7	Realizan muy buenas exhibiciones de productos.					
V8	El área de compra es espaciosa.					
V9	Las mercancías están muy bien colocadas.					
V10	El pago se realiza con rapidez.					
V11	El surtido de productos es suficiente para mis necesidades.					
V12	Realiza publicidad frecuente (en folletos, prensa local...) de sus productos.					

Los diferentes resultados obtenidos permitieron elaborar perfiles de percepción e imagen de los detallistas por los consumidores, desglosando la información con respecto a las siguientes variables: zonas geográficas, segmentos de compradores, fórmula de venta adoptada por los establecimientos minoristas, cada economato, hipermercado y cadena comercial de superservicios-supermercados y tabulaciones cruzadas de estas variables. Comentamos brevemente las conclusiones más significativas, principalmente las relativas a cada tipo de detallista y tipologías de consumidores según la edad.

Por término medio, la valoración para la mayoría de los atributos considerados es más positiva a medida que se incrementa la superficie de venta del detallista (véase Gráfico 3.L.5). Efectivamente, de mayor a menor predisposición hacia la compra podemos citar: hipermercado, superservicio/supermercado, autoservicio y tienda tradicional. Apreciamos dos resultados que no son acordes con la afirmación previa:

1. La tienda tradicional es donde el pago de la mercancía comprada se realiza con mayor celeridad y donde el personal de ventas es más amable y solícito.
2. Para determinados productos, el hipermercado se percibe con precios superiores a los del superservicio/supermercado.

El economato es considerado como el establecimiento con menores puntuaciones en los siguientes atributos: tienda agradable, limpio y cuidado, personal amable y solícito, colocación de la mercancía y publicidad realizada. Sin embargo, cuenta con los precios más baratos.

A medida que aumenta la edad del entrevistado, su opinión sobre los distintos tipos de detallistas, en cada uno de los atributos investigados, es más favorable. Excepciones a este comentario son las personas de más de 60 años, quienes no consideran muy atractivas las compras en el hipermercado.

III.2. Imagen y posicionamiento competitivo: Aplicaciones del análisis factorial

En este epígrafe estudiamos la imagen proyectada por los establecimientos detallistas, posicionándolos en relación a sus competidores. Para ello, se utilizó la es- caía de actitud correspondiente a una de las preguntas del cuestionario (véase Cuadro 3.L.1). El tratamiento de la información obtenida

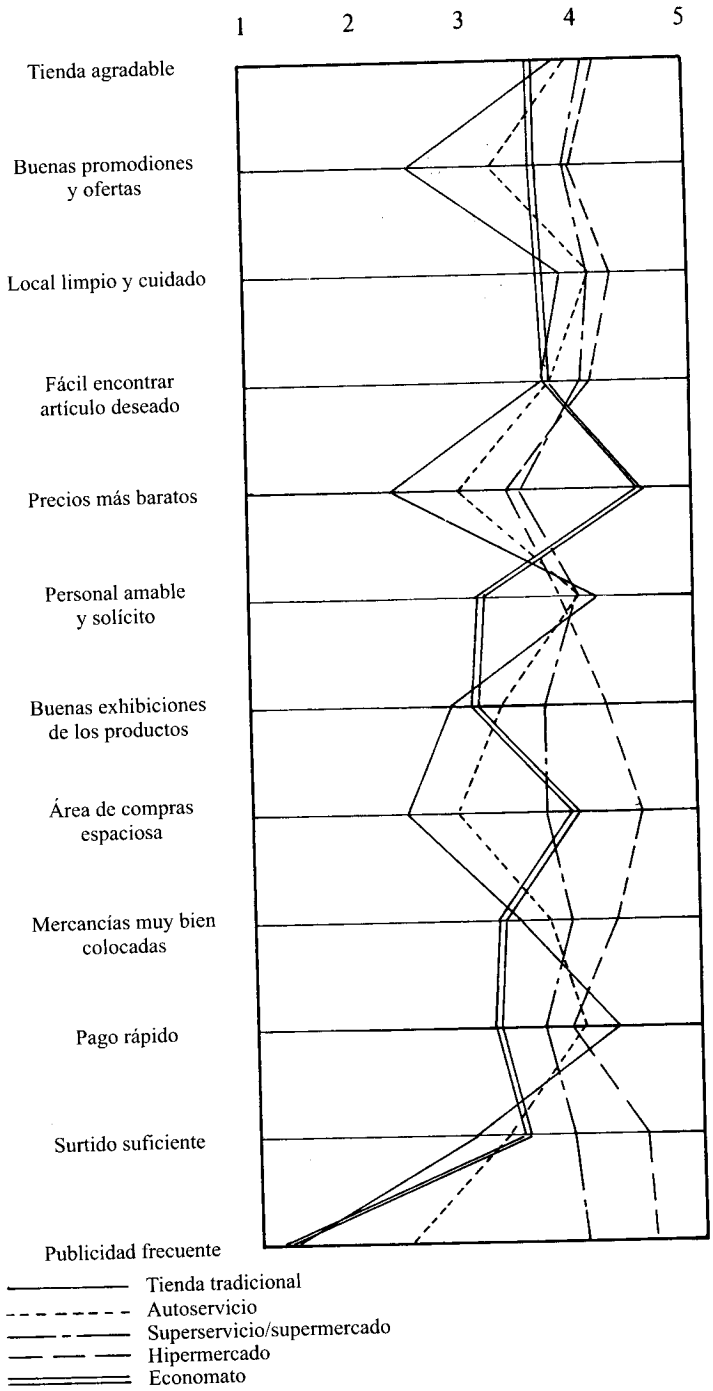


GRÁFICO 3.L.5. Perfil de imagen para los establecimientos detallistas.

mediante dicha escala debe ser objeto de comentarios más exhaustivos. Afrontamos, a continuación, esta tarea explicativa.

Para un mercado dado, los establecimientos detallistas en competencia pueden ser percibidos de una forma muy diferente en función de un conjunto de atributos. Estudiar la posición relativa de cada detallista viene a ser, pues, intentar situarlo en su contexto competitivo, desgajando los elementos constitutivos de la percepción comparativa que lo consumidores tienen de dicho establecimiento. Dado que e número de atributos iniciales, en función de los que se va loraban los detallistas, puede resultar no manejable de car al posicionamiento en un contexto competitivo, es oportu no proceder a su reducción.

Efectivamente, parece aconsejable emplear un instru mento de cálculo cuya función se limite a la transformació

de un conjunto de atributos de tamaño normalmente elevado, en otro conjunto de dimensión mínima con la menor pérdida de información. Estos nuevos atributos los denominaremos *factores* y la técnica para obtenerlos *análisis factorial* cuyo tratamiento en computador se conoce como *análisis de componentes principales con rotación VARIMAX* Comentarios adicionales al respecto facilitarán al lector la interpretación de la imagen y actitud de los detallistas posicionándolos en el mercado a estudiar:

1. Recordar que cada factor es una combinación de los atributos iniciales con los que ha valorado el entrevistado a los detallistas. Por tanto, su interpretación hay que buscarla como combinación de los significados de las variables con las que se relacionan. Cuanto más participe un atributo original en la formación de un factor más intervendrá en el significado que a este último se asocia.

   Tampoco del nombre que subjetivamente se asigne a cada factor hay que fiarse demasiado puesto que en ocasiones la tarea es difícil. En realidad se trata de un título que sirva para referirse al factor con un significado que, en cierto modo, nos haga recordar los atributos que explicaba. Queremos insistir sobre este punto, porque es frecuente que se produzca un fenómeno de «apego al título» que puede llevar a desviaciones en la interpretación. No se debe olvidar nunca las connotaciones o atributos en los que se fundamenta cada factor, quedándose tan sólo a nivel del título y asimilando otras interpretaciones que éste pueda tener y que a veces son diferentes de las derivadas de los atributos que más intervienen en formar cada factor.
2. Intuitivamente la sustitución de unos atributos iniciales por unas variables resumen responde a una sencilla y breve interpretación. El consumidor en su estructura mental no es capaz de separar una serie de atributos. La tenencia de un atributo implica instantáneamente la posesión de otro. A priori, esto es difícil de detectar por el investigador. Sin embargo, la obtención a posteriori de factores (utilizando como información la respuesta a la escala de actitud propuesta) ofrece una amplia visión de esa estructura latente. La determinación de factores guiará la política de comunicación empresarial y el posicionamiento de cada detallista en comparación con los restantes del mercado estudiado.

En resumen, nuestro objetivo es doble: eliminar información redundante para *detectar estructuras latentes* del consumidor cuando realiza compras en establecimientos minoristas y utilizar los factores obtenidos para *construir espacios bidimensionales* y posicionar en los mismos todos los detallistas analizados, comentando zonas de gran competencia, similitud de opiniones, actitud ante los detallistas y delimitación e investigación de posibles huecos de mercado.

Subrayar que se llevaron a cabo diversas aplicaciones de la técnica de análisis factorial: para el conjunto de encuestas realizadas en todos los establecimientos detallistas; para las opiniones de los entrevistados en cada tipo de minorista; para diferentes segmentos de consumidores.

El tratamiento de las respuestas de toda la muestra de clientes mediante análisis factorial puso de manifiesto la existencia de tres factores significativos que resumían los aspectos racionales y motivacionales integrados en los doce atributos iniciales. Detallamos el contenido de estos factores para después utilizarlos en un posicionamiento gráfico de los detallistas (8).

Recordar que el significado de un factor hay que buscarlo en los atributos de partida que han servido para construirlo. La relación entra cada factor y los atributos iniciales viene reflejada por la «matriz de cargas factoriales» (cuanto mayores sean estos conceptos en valor absoluto, más importancia tendrán los respectivos atributos en explicar cada factor). De acuerdo con este comentario (véase Cuadro 3.L.2), el factor 1 se relaciona con el conjunto de técnicas de venta cuyo objetivo es mejorar la presentación del producto para así aumentar su rotación en los establecimientos detallistas e incidir positivamente sobre el beneficio.

CUADRO 3.L.2. Atributos que más intervienen en los factores obtenidos estudio imagen establecimientos detallistas

Atributos	*Factor 1*	*Factor 2*	*Factor 3*
V8 8	0,766	0,000	0,000
V9 9	0,700	0,372	0,000
V7 7	0,684	0,275	0,000
V11 11	0,608	0,000	0,293
V6 6	0,000	0,825	0,000
V1 1	0,000	0,723	0,294
V10 10	0,000	0,597	–0,322
V3 3	0,348	0,534	0,000
V2 2	0,000	0,000	0,837
V5 5	0,000	0,000	0,798
V4 4	0,407	0,490	0,000
V12 12	0,497	0,000	0,470
Varianza explicada acumulada	0,557	0,838	1,000

(8) El programa utilizado es el BMDP-4M. Salvo ligeras diferencia estos tres factores también se hacen patentes cuando el análisis factorial se aplica individualmente a los clientes de cada tipo de detallista. La única diferencia es que para los clientes del hipermercado, el factor que engloba aspectos de la politice de merchandising se desdobla en dos: a) facilidad para realizar las compras (fácil de encontrar el articulo deseado debido a pasillos amplios, vitrinas adecuadas, buena señalización de precios y de secciones de venta; mercancía bien colocada); b) surtido/atractividad (incitación a la compra mediante publicidad frecuente, incluso en el punto de venta, que dé a conocer la amplitud y extensión del surtido; exhibición de los productos de forma adecuada, permitiendo que el surtido sea atractivo para el cliente y se venda por sí solo).

Por un lado, alude al tamaño del punto de venta: área de compras espaciosa. Ello permite un surtido más adecuado (en relación a las líneas de productos, familias o marcas) y un abanico de alternativas para colocar la mercancía en secciones y lineales. Así, el comprador puede disfrutar con el acto de ir de compras beneficiándose de una adecuada exhibición de los productos. Un cierto número de entrevistados opinan que los minoristas de mayor tamaño invierten más en publicidad para fomentar y desarrollar su imagen; disponen de un gran conjunto de medios (físicos, psicológicos y de estimulo) que hacen atractiva la compra, incitan a la misma y facilitan encontrar los artículos deseados. Denominamos al factor 1 como *percepción de la política de merchadising del detallista.*

El factor 2 representa aspectos relacionados con la política de personal del detallista. Las tiendas cuyo personal es amable y solícito (en ocasiones indicando dónde encontrar el artículo deseado, entregándolo personalmente) son más agradables para comprar. Incluso, una buena política de personal reduce la espera para el pago en caja y permite tener el establecimiento más limpio y cuidado. Denominaremos al factor 2 como de *tratamiento y servicio personal.*

Finalmente, el factor 3 es susceptible de una interpretación muy sencilla, dado que engloba aspectos económicos: los establecimientos con precios más baratos son aquellos que implantan destacadas promociones y ofertas, aunque otras muchas veces deben ser acompañadas por publicidad (por ejemplo por correo, prensa u otro medio de comunicación) e incluso marketing directo para poder ser percibidas. Por tanto, conceptualizamos el factor 3 como de *economía.*

Un resumen de lo expuesto se representa en los diversos apartados del Gráfico 3.L.6, donde los atributos iniciales son vectores, tanto más próximos a cada factor cuanto más intervengan en su explicación.

Los factores comentados también son útiles para posicionar tanto tipos de detallistas como empresas concretas que actúan en dicho mercado. Parte de estos resultados se pueden apreciar en los Gráficos 3.L.7, 3.L.8 y 3.L.9 de los que se pueden deducir conclusiones en cuanto a:

1. La imagen comparativa de los competidores en el mercado al por menor. La tienda tradicional y el autoservicio se caracterizan por un mejor tratamiento y servicio personal. Los superservicios/supermercados son percibidos como de mayor economía que los hipermercados, mientras que estos últimos obtienen ventajas competitivas mediante una eficiente política de merchandising.
2. El factor más determinante (y por derivación los atributos que cada uno explica) en función de la variación que en los mismos presentan las empresas minoristas. Efectivamente, para un factor concreto cuanto mayor sea la diferencia entre las puntuaciones factoriales de los detallistas, más importancia se le asignará ya que puede ser objeto de posibles estrategias de diferenciación (en este caso política de merchandising y economía).

III.3. Imagen y motivaciones de compra

La modificación de la imagen de las empresas detallistas mediante la formulación y desarrollo de estrategias competitivas no sólo debe considerar las conclusiones expuestas en los epígrafes previos sino también en qué medida éstas son acordes con las motivaciones de compra del consumidor. Teniendo en cuenta estas afirmaciones, ofrecemos diversos resultados que pretenden ayudar al lector interesado a descubrir y explotar oportunidades en el mercado estudiado (véase Cuadro 3.L.3).

CUADRO 3.L.3. Motivaciones de compra

Motivos compra (1)	*Total*	*Alimentación bebidas*	*Carne charcutería*	*Fruta verdura*	*Droguería limpieza*	*Perfumería*
Limpieza	38,54	32,70	73,00	64,80	12,30	10,20
Precios	75,50	80,20	63,50	69,90	85,10	78,70
Atención al cliente	34,98	27,10	39,40	31,70	24,70	52,20
Surtido calidad (2)	74,81	66,30	90,70	86,20	56,70	74,40
Promociones y ofertas	40,28	50,10	15,40	22,20	68,00	45,30
Proximidad al comicilio	35,34	43,7	35,90	32,90	36,80	27,20
Rapidez en el servicio	19,45	20,00	22,50	17,10	21,10	16,50
Servicio a domicilio	2,95	3,80	2,80	3,30	3,20	1,60

(1) Principales porcentajes de respuesta sobre el total de entrevistados.
(2) Se recogen componentes físico-químicos y de notoriedad de marca para aseverar que los productos son de calidad.

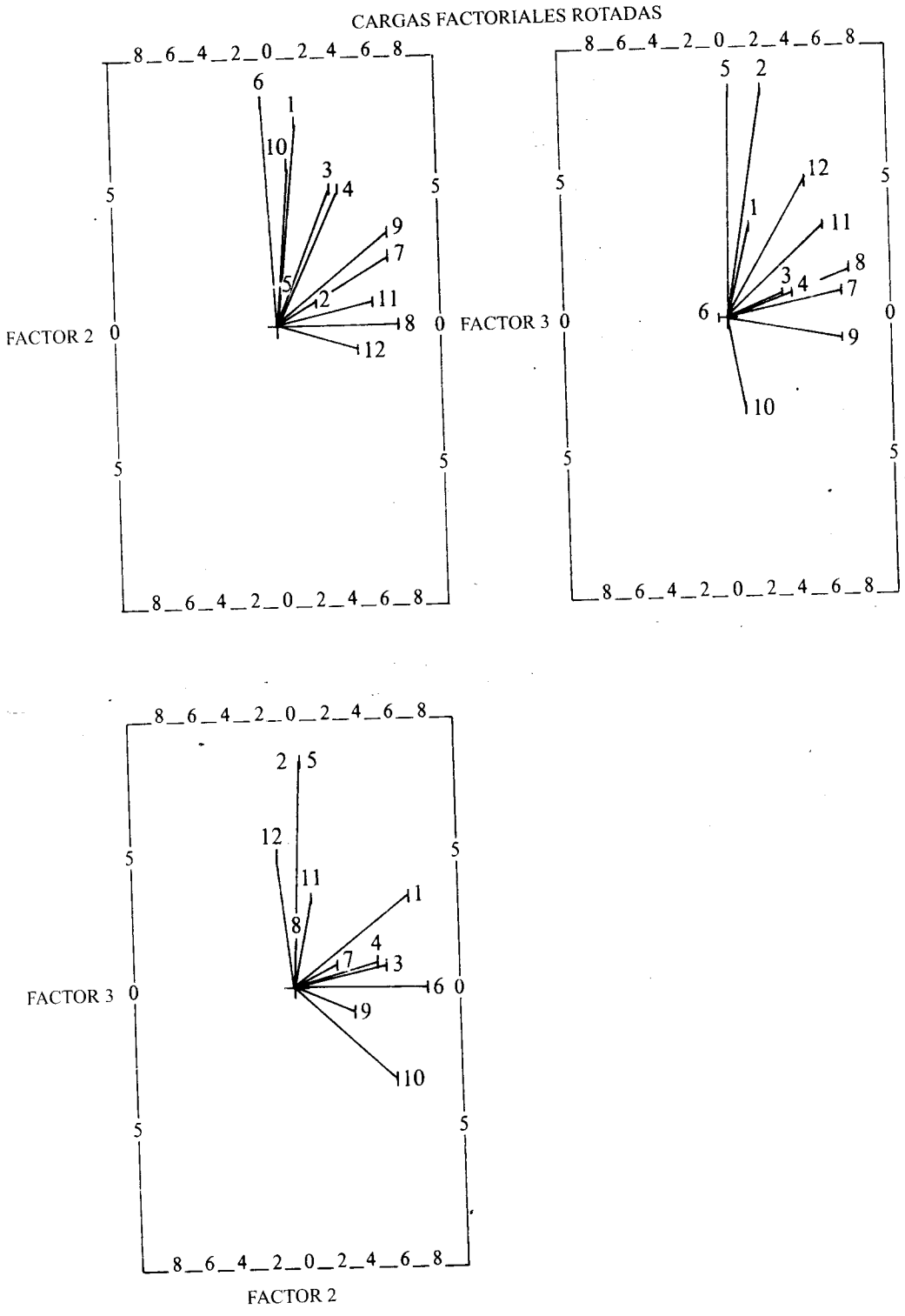


GRÁFICO 3.L.6. Posicionamiento atributos.

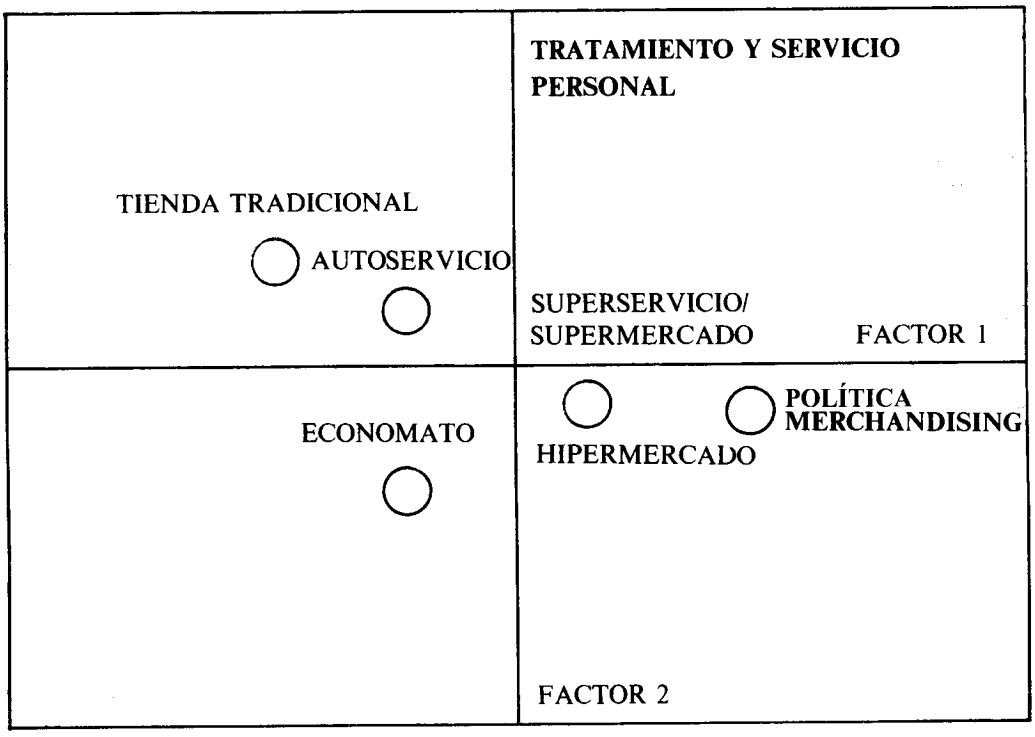


GRÁFICO 3.L.7.

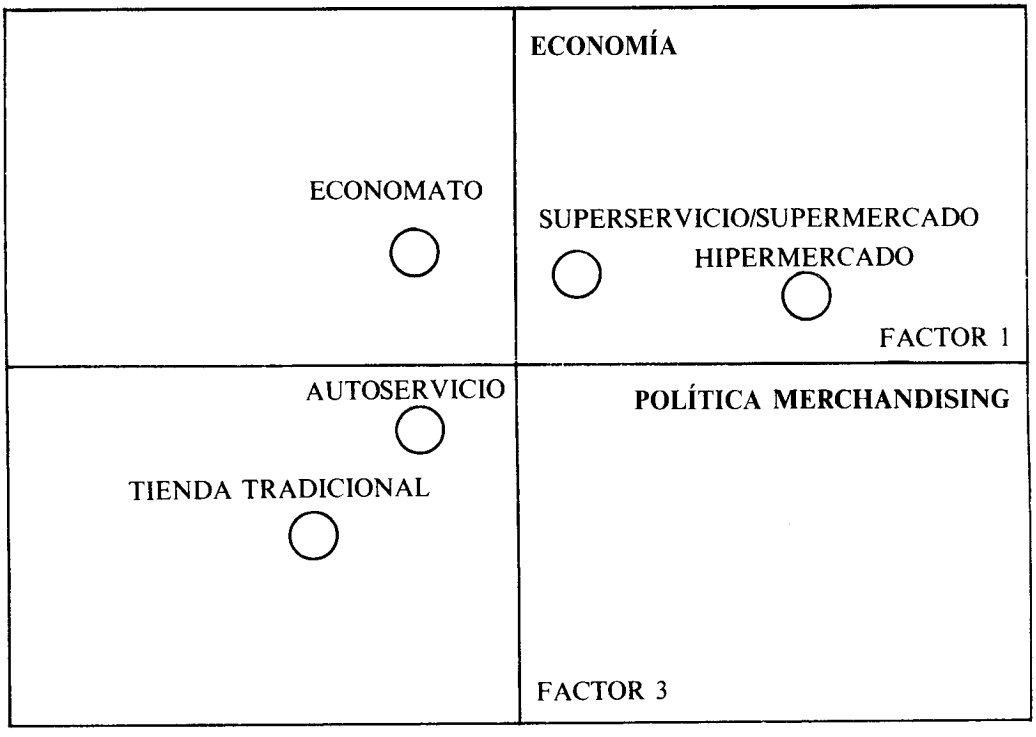


GRÁFICO 3.L.8.

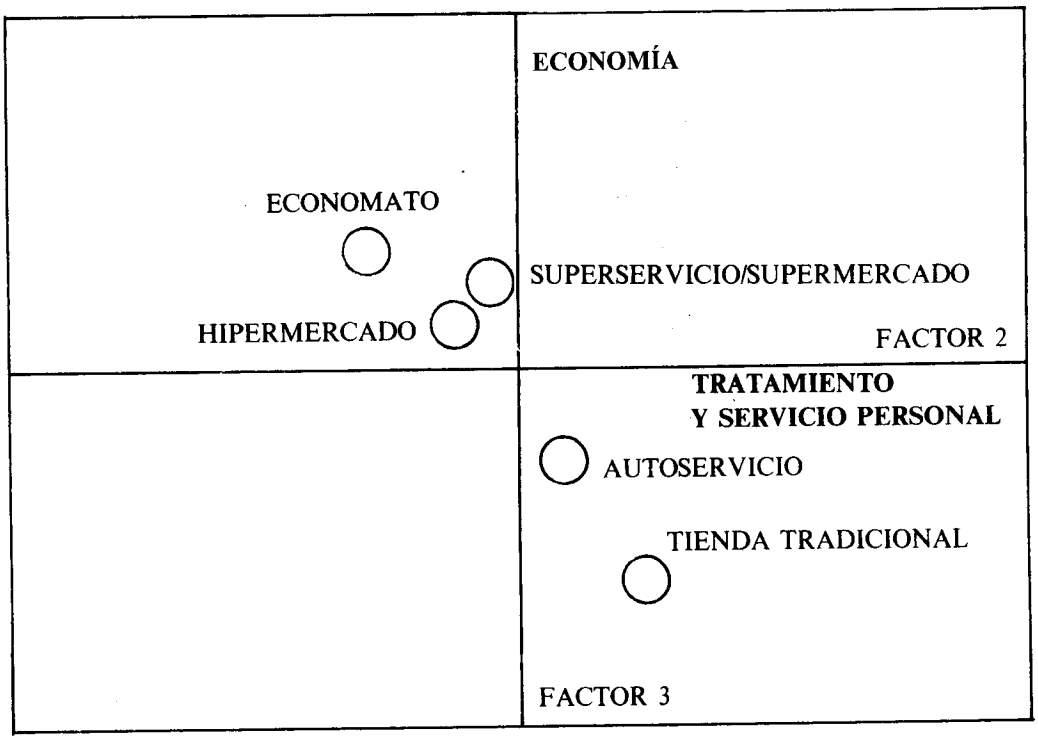


GRÁFICO 3.L.9.

IV. CONCLUSIONES

Las conclusiones expuestas en este artículo tienen un doble objetivo: presentar una breve panorámica de las principales investigaciones relativas a los aspectos que configuran la imagen de los establecimientos detallistas y desarrollar un estudio empírico tendente a analizar la imagen proyectada por los minoristas de alimentación (en una zona geográfica concreta) o percepción del servicio compuesto por las diferentes utilidades que de su actividad se deriva para el consumidor. Se estiman perfiles de imagen, mapas de percepción mediante la técnica de análisis factorial y motivaciones de compra para distintos tipos de detallistas.

De cualquier forma, téngase en cuenta que los resultados comentados tan sólo pueden tomarse como indicativos de la imagen en el mercado. Somos conscientes de que el estudio puede incorporar modificaciones adicionales que, sin duda, redundarían en beneficio de las conclusiones obtenidas. Particularmente en cuanto a:

1. Utilización de un mayor número de atributos para una más exacta cuantificación del constructo imagen.
2. Posibilidad de tratamiento de la información mediante metodologías adicionales, como el análisis factorial de correspondencias, con el fin de facilitar la interpretación de los factores.
3. Enfatizar la relación imagen-preferencias del consumidor.
4. Investigar con mayor profusión la imagen percibida por diferentes grupos de clientes considerando no sólo variables sociodemográficas (edad, número de miembros de la familia y ocupación profesional), sino también estilos de vida y orientación hacia la compra.
5. Análisis de la evolución de la imagen de las diferentes empresas detallistas en el tiempo.

V. BIBLIOGRAFÍA

ANDRES, A. J.: *Apuntes de Merchandising.* Góndola. Editmex, 1983. ARNOLD, D. R.; CAPELLA, L. M. y SMITH, G. D.: *Strategic Retail Management.* Addison-Wesley Publishing Company, Inc., 1983.

BELLENGER, D. N. y KORGAONKAR, P. K.: *Profiling the Recreational Shopper.* Journal of Retailing, 62 (otoño), pp. 77-92, 1986.

BLACK, W. y REILLY, M.: *A Taxonomy of Shopping Trips Incorporating Multistop and Multipurpose Behavior.* Working Paper Series. Tucson, Ariz.: University of Arizona, College of Business Administration, 1983.

DELLA BRITTA, A. J. y MONROE, K. B.: *A Multivariate Analysis of the Percpetion of Value for Retail Price Advertisements.* Advances in Consumer Research, 8 (Ann. Arbor, Mich. Association for Consumer Research), pp. 161-165, 1981.

DELLA BRITTA, A. J.; MONROE, K. B. y McGINNIS, J. M.: *Consumer Perceptions of Comparative Price Advertisements.* Journal of Marketing Research, vol. 18 (noviembre), pp. 416-427, 1981.

DIJKSTRA TAU REL, M.: *El Reto de las Marcas Propias.* Distribución Actualidad, octubre, pp. 151-161, 1987.

DONOVAN, R. J. y ROSSITER, J. R.: *Store Atmosphere: An Enviromental Psychology Approach.* Journal of Retailing, 58 (primavera), pp. 34-576, 1982.

DUNCAN, D.; HOLLANDER, 5. y SAVITT, R.: *Modern RetailingManagemenL* Homewood, 111.: Richard D. Irwin, Inc., 1983.

GUILTINAN, J. P. y MONROE, K. B.: *Identifyinging and Analyzing Consumer Shop ping Strategies,* en OLSON, J. (ed.): *Advances in Consumer Research.* (Chicago: Association for Consumer Research), pp. 745-748, 1980.

HARRIS, B. F. y STRANG, R.: *Marketing Strategies in the Age of Generics.* Journal of Marketing, vol. 49 (otoño), pp. 70-81, 1985.

HAWKINS, D. I.; BEST, R. J. y CONEY, K. A.: *Consumer Behavior: Implications for Marketing Strategy.* Business Publications, Inc. pp. 648-654, 1986.

HIGIE, R. A.; FEICK, L. F. y PRICE, L. L.: *Types and Amount of Word-of-Mouth Communications About Retailers.* Journal of Retailing, 63 (otoño), pp. 260-278, 1987.

HIRSCHMANN, E. C.; GREENBERG, E. y ROBERSTON, D. H.: *The Intermarket Reliability of Retail Image Research: An Empirical Examination.* Journal of Retailing, 54 (primavera), pp. 3-12, 1978.

HOWES, J. M. y LUMPKIN, J. R.: *Understanding the Outshopper.* Journal of the Academy of Marketing Science, 12 (otoño), pp. 200-216, 1984.

HUMMEL, J. W. y SAVITT, R.: *Integrated Customer Service and Retail Strategy.* International Journal of Retailing, vol. 3, 2, pp. 5-21, 1988.

JAMES, D. L.; DURAND, R. y DREVES, R. A.: *The Use of a Multi Attribute Attitude Model in a Store Image Study.* Journal of Retailing, 58 (verano), pp. 64-86, 1976.

KUMAR, V. y LEONE, R. P.: *Measuring the Effect of Retail Store Promotions on Brand and Store Substitution.* Journal of Marketing Research, vol. 25 (mayo), pp. 178-185, 1988.

LEWISON, D. y DELOZIER, W.: *Retalling.* Columbus, Ohio: Merril Publishing Company, 1986.

MALHOTRA, N. K.: *A Threshold Model of Store Choice.* Journal of Retailing, 59 (verano), pp. 3-21, 1983.

MASSON, J. E. y WELLHOFF, A.: *El Merchandising. Rentabilidad y Gestión en el Punto de Venta.* Ediciones Deusto. Bilbao, 1984.

MAZURSKY, D. y JACOBY, J.: *Exploring the Development of Store Images.* Journal of Retailing, 62 (verano), pp. 145-165, 1986.

McENALLY, M. R. y HAWES, J. M.: *The Market for Generic Brand Grocery Products: A Review and Extension.* Journal of Marketing, vol. 48 (invierno), pp. 75-83, 1984.

OLSHAVSKY, R. y CRASK, M.: *The Multi-Item, Multi-Stop Store Choice Process: An Exploratory Study,* en BAGOZZI, R. P. y TYBOUT, A. (eds.): *Advances in Consumer Research,* vol. X, pp. 351-355, 1983.

PESSEMIER, E. A.: *Store Image and Positioning.* Journal of Retailing, 56 (primavera), pp. 94-107, 1980.

RUNYON, K. E. y STEWART, D. W.: *Consumer Behavior.* Merril Publishing Company. Tercera Edición. Cap. 17, 1987.

SALEN, H.: *Distribución y Merchandising. Factores Claves del Éxito.* Ed. Salen, H. 1987.

SALEN, H.: *Distribución: Cómo Ser el Número 1.* Ed. Salen, H. 1988.

SCHIFFMAN, L. G.; DASH, J. F. y DILLON, W. R.: *The Contribution of Store Image Characteristics to Store-Type Choice.* Journal of Retailing, 53 (verano), pp. 3-14, 1977.

SHARMA, S. y MALHOTRA, N. K.: *Examining Social Setting and Product Preference via Conjoint Analysis: An Empirical Study of Restaurant Patronage,* en BELLUR, V. V.; *Developments in Marketing Science,* 4. Academy of Marketing Science, pp. 59-63, 1981.

SINGSON, R. L.: *Multidimensional Scaling Analysis of Store Image and Shopping Behavior.* Journal of Retailing, 51 (verano), pp. 38-52, 1975.

SPIGGLE, S. y SEWALL, M. A.: *A Choice Sets Model of Retail Selection.* Journal of Marketing, vol. 51 (abril), pp. 97-111, 1987.

VÁZQUEZ CASIELLES, R. y TRESPALACIOS, J. A.: *La Selección de Establecimientos Detallistas por el Consumidor: Análisis de Imagen y Posicionamiento Competitivo.* Comercialización e Investigación de Mercados. Universidad de Oviedo, pp. 1-160, 1989.

WALTERS, R. G.: *Retail Promotions and Retail Store Performance: A Test of Some Key Hypotheses.* Journal of Retailing, 64 (verano), pp. 153-180, 1988.

WALTERS, R. G. y RINNE, H.: *An Empirical Investigation into the Impact of Price Promotions on Retail Store Performance.* Journal of Retailing, 62 (otoño), pp. 237-266, 1986.

WESTBROOK, R. A.: *Sources of Consumer Satisfaction with Retail Outlets.* Joumal of Retailing, 57 (otoño), pp. 68-85, 1981.

WESTBROOK, R. A. y BLACK, W. C.: *A Motivation-Based Shopper Typology.* Journal of Retailing, 61 (primavera), pp. 78-103, 1985.

WILKINSON, J. B.; MASON, J. B. y PAKSOY, C.: *Assessing the Impact of Short-Term Supermarket Strategy Variables.* Journal of Marketing Research, vol. 19 (febrero), pp. 72-86, 1982.

SECCIÓN

2

TÉCNICAS DE DEPENDENCIA

Consideración general

Mientras que la Sección 1 se centró en la preparación de datos para el análisis multivariante, la Sección 2 aborda lo que muchos denominarían la esencia del análisis multivariante: las técnicas de dependencia. Tal y como se señaló en el Capítulo 1, las técnicas de dependencia se basan en el uso de una serie de variables independientes para predecir y explicar una o más variables dependientes. El investigador, tanto si se enfrenta a variables dependientes de naturaleza métrica o no métrica, tiene a su disposición una variedad de métodos de dependencia que le ayudan en el proceso de relacionar las variables independientes con las variables dependientes. Dada la naturaleza multivariante de estos métodos, todas las técnicas de dependencia tienen en cuenta las variables independientes múltiples mientras que permiten las variables dependientes múltiples en ciertas situaciones. Por tanto, el investigador tiene a su disposición una serie de técnicas que en principio dan cabida a las necesidades de análisis de casi cualquier tipo de materia de investigación que tenga que ver con una relación de dependencia. Además, estas técnicas proporcionan la oportunidad no solamente de una mayor capacidad de predicción sino también de una mejor explicación de la relación entre la variable criterio con las variables independientes. La explicación es cada vez más importante conforme la materia de investigación empieza a abordar las cuestiones acerca de cómo funciona la relación entre variables independientes y dependientes.

Capítulos de la Sección 2

La Sección 2 cubre seis técnicas de dependencia: la regresión múltiple, el análisis discriminante, el análisis multivariante de la varianza, la regresión logística, el análisis conjunto y la correlación canónica, desde los Capítulos 4 a 8. Las técnicas de dependencia, tal y como comentamos anteriormente, permiten al investigador valorar el grado de relación entre variables dependientes e independientes. Dichas técnicas varían en cuanto al tipo y el carácter de la relación, tal y como se refleja en las

propiedades de medición de las variables dependientes e independientes abordadas en el Capítulo 1. Por ejemplo, tanto la regresión múltiple, como el análisis discriminante o la correlación canónica contienen variables independientes métricas múltiples, que sin embargo varían en el tipo de variable criterio (análisis de regresión —métrica simple; análisis discriminante —no métrica simple; y correlación canónica —métrica múltiple). El Capítulo 4, análisis de regresión múltiple, se centra en lo que es quizá la más fundamental de todas las técnicas multivariantes, al tiempo que representa una «base esencial» para nuestra discusión sobre otros métodos de dependencia. Tanto si se trata de la valoración del cumplimiento de los supuestos estadísticos subyacentes, como si se trata de la interpretación del valor teórico de las variables independientes, se considerará que las cuestiones abordadas en el Capítulo 4 son tan cruciales como lo resultan en muchas de las otras técnicas. El Capítulo 5, análisis discriminante múltiple y regresión múltiple, investiga una forma de relación de dependencia única —una variable criterio no métrica. En esta situación, el investigador intenta clasificar las observaciones en grupos. Esta clasificación en grupos puede alcanzarse mediante el análisis discriminante o la regresión logística, una variante de la regresión diseñada para tratar específicamente con las variables dependientes no métricas. En el Capítulo 6, análisis multivariante de la varianza, el análisis difiere en varios aspectos al de las técnicas anteriores, dado que se adecúa al análisis de variables dependientes métricas y variables independientes no métricas.

Aunque esta técnica es una extensión directa del análisis simple de la varianza, las variables dependientes métricas múltiples hacen que la predicción y la explicación resulten más complejas. El Capítulo 7, análisis conjunto, nos presenta una técnica que no se asemeja a ninguno de los otros métodos multivariantes anteriormente explicados, dado que el investigador determina los valores de las variables no métricas independientes en una manera casi experimental. Una vez diseñado el estudio, el encuestado proporciona únicamente información acerca de la variable criterio. Aunque representa más responsabilidad para el investigador, el análisis conjunto proporciona una herramienta poderosa para entender procesos de decisión complejos. Finalmente, en el Capítulo 8, correlación canónica, el investigador se enfrenta a la forma de análisis multivariante más generalizada, que utiliza las múltiples variables dependientes e independientes. En situaciones en que existen valores teóricos tanto para variables dependientes como independientes, la correlación canónica proporciona un método flexible de predicción y explicación.

Esta sección proporciona al investigador una exposición de la amplia gama de técnicas de dependencia disponible, cada una de ellas adecuada a una tarea y relación específica. Cuando se complete, los problemas acerca de la selección de estos métodos deberían ser de fácil solución, resultando asimismo asequible al lector la interpretación de sus resultados.

CAPÍTULO 4

Análisis de regresión múltiple

OBJETIVOS DE APRENDIZAJE

- Diferenciar las técnicas de análisis de regresión de otras técnicas multivariantes.
- Determinar cuándo el análisis de regresión es la herramienta estadística apropiada para analizar un problema.
- Entender cómo nos puede ayudar la regresión para hacer predicciones usando el concepto de mínimos cuadrados.
- Ser consciente de los importantes supuestos que subyacen en el análisis de la regresión y estar preparado para instrumentar las soluciones cuando se incumplen tales supuestos.
- Interpretar los resultados de la regresión tanto desde un punto de vista estadístico como empresarial.
- Aplicar los procedimientos de diagnóstico necesarios para evaluar las observaciones influyentes.
- Explicar las diferencias entre las regresiones simultáneas y por pasos.
- Usar las variables ficticias y comprender su interpretación.

Este capítulo describe el análisis de regresión múltiple tal y como se emplea para resolver importantes problemas de investigación, particularmente en el mundo empresarial. El análisis de regresión es con mucho la técnica de dependencia más versátil y ampliamente utilizada, aplicable en cualquier ámbito de la toma de decisiones en los negocios. Sus usos van de los problemas más generales a los más específicos, relacionando en cada caso un factor (o factores) con un resultado específico. Por ejemplo, el análisis de regresión es el fundamento de los modelos de previsión económica, que van desde los modelos econométricos de predicción de la economía nacional ba-

sados en ciertos inputs (niveles de renta e inversión) a los modelos de comportamiento de una empresa en el mercado si se sigue una estrategia de marketing determinada. Los modelos de regresión se utilizan también para estudiar cómo los consumidores toman decisiones o forman impresiones o actitudes. Otras aplicaciones incluyen la evaluación de los determinantes de la efectividad de un programa (por ejemplo, qué factores ayudan a mantener la calidad) y de la viabilidad de un nuevo producto o el rendimiento esperado de una nueva acción que cotiza en bolsa. Incluso, aunque estos ejemplos ilustren sólo un pequeño subconjunto de todas las aplicaciones, demuestran que el análisis de regresión es una poderosa herramienta analítica diseñada para explorar todos los tipos de relaciones de dependencia.

El análisis de regresión múltiple es una técnica estadística general utilizada para analizar las relaciones entre una única variable criterio y varias variables independientes. Como se señaló en el Capítulo 1, su formulación básica es:

$$\underset{\text{(Métrica)}}{Y_1} = \underset{\text{(Métrica)}}{X_1 + X_2 + \cdots + X_n}$$

Este capítulo presenta las líneas generales para juzgar la conveniencia de la regresión múltiple respecto a varios tipos de problemas. Se van a proporcionar sugerencias para interpretar los resultados de su aplicación tanto desde un punto de vista estadístico como de toma de decisiones. Se examinarán las posibles transformaciones de los datos para remediar las violaciones de los diversos supuestos, junto con una serie de procedimientos de diagnóstico que identifican observaciones con una influencia particular sobre los resultados. Muchos lectores que ya tengan conocimientos previos de los procedimientos de regresión múltiple pueden obviar la primera parte del capítulo. Pero para aquellos que estén menos familiarizados con la materia, este capítulo proporciona un valioso fundamento para el estudio del análisis multivariante de los datos.

¿Qué es el análisis de regresión múltiple?

El análisis de regresión múltiple es una técnica estadística que puede utilizarse para analizar la relación entre una única **variable criterio (criterio)** y varias **variables independientes (predictores)**. El objetivo del análisis de regresión múltiple es usar las variables independientes cuyos valores son conocidos para predecir la única variable criterio seleccionada por el investigador. Cada variable predictor es ponderada, de forma que las ponderaciones indican su contribución relativa a la predicción conjunta. Al calcular las ponderaciones, el procedimiento del análisis de regresión asegura la máxima predicción a partir del conjunto de variables independientes. Estas ponderaciones facilitan también la interpretación de la influencia de cada variable en la realización de la predicción, aunque la correlación entre las variables independientes complica el proceso de interpretación. El conjunto de variables independientes ponderadas es conocido también como **valor teórico de la regresión**, una combinación lineal de las variables independientes que predice mejor la variable criterio (el Capítulo 1 contiene una explicación más detallada del valor teórico). La ecuación de regresión, también denominada como el valor teórico de la regresión, es el ejemplo de valor teórico más ampliamente reconocido entre todas las técnicas multivariantes.

Como ya se mencionó en el Capítulo 1, el análisis de regresión múltiple es una técnica de dependencia. Por tanto, al utilizarla, deberíamos de ser capaces de dividir las variables entre independientes y dependientes. El análisis de regresión es también una herramienta estadística que debería utilizarse sólo cuando tanto las variables dependientes como las independientes son métricas. Sin embargo, bajo ciertas circunstancias, es posible incluir datos no métricos para las variables independientes (transformando los datos ordinales o los nominales en variables ficticias) o la variable criterio (mediante el uso de una medida binaria en la técnica especial de la regresión logística). En resumen, al aplicar el análisis de regresión múltiple: (1) los datos deben ser métricos o apropiadamente transformados y (2) antes de derivar la ecuación de regresión, el investigador debe decidir qué variable va a ser dependiente y cuál de las restantes variables será independiente.

Un ejemplo de la regresión simple y múltiple

El objetivo del análisis de regresión es predecir una única variable criterio a partir del conocimiento de una o más variables independientes. Cuando el problema implica una única variable independiente, la técnica estadística se denomina **regresión simple**. Cuando el problema implica dos o más variables independientes, se denomina **regresión múltiple**. Para ilustrar los principios básicos relacionados con la utilización de esta técnica, se proporcionan los resultados de un pequeño estudio de ocho familias y su uso de tarjetas de crédito. Se identificaron tres factores potenciales (tamaño familiar, ingresos familiares y el número de automóviles poseídos), y se recogieron datos de cada una de las ocho familias (véase Tabla 4.1). En la terminología del análisis de regresión, la variable criterio (Y) es el número de tarjetas de crédito utilizado y las tres variables (V_1, V_2 y V_3) representan el tamaño de familia, los ingresos familiares y el número de automóviles poseídos, respectivamente. La exposición de este ejemplo se divide en tres partes para ayudar a entender cómo la regresión estima la relación entre la variable independiente y la variable criterio. Los tres temas que se van a tratar son (1) predicción sin una variable independiente, utilizando sólo una medida única —la media, (2) la predicción usando una única variable independiente —regresión simple, y (3) predicción usando varias variables independientes —regresión múltiple.

TABLA 4.1. Resultados de la encuesta sobre el uso de tarjetas de crédito

Familia ID	*Número de posesión de tarjetas de crédito* Y	*Tamaño de familia* V_1	*Renta familiar (000$)* V_2	*Número de posesión de automóviles* V_3
1	4	2	14	1
2	6	2	16	2
3	6	4	14	2
4	7	4	17	1
5	8	5	18	3
6	7	5	21	2
7	8	6	17	1
8	10	6	25	2

Fijando una línea de base: predicción sin variable independiente

Antes de hacer la estimación con la primera ecuación de regresión, empecemos con el cálculo de una línea básica con la que compararemos la capacidad de predicción de nuestros modelos de regresión. La linea básica debería representar nuestra mejor predicción sin el uso de variables independientes. Podríamos usar cualquier número de opciones (por ejemplo, la predicción perfecta, un valor especificado previamente o una de las medidas de tendencia central, como la media, la mediana o la moda), sin embargo la línea predictor utilizada en la regresión es la media simple de la variable dependiente, lo cual tiene varias propiedades deseables. En nuestro ejemplo, la media aritmética del número de tarjetas utilizadas es siete. Nuestra predicción podría ser «el número medio de tarjetas de crédito mantenidas por una familia es siete». También podríamos poner esto como la siguiente ecuación de regresión:

Predicción del número de tarjetas de crédito = Número medio de tarjetas de crédito

o

$$\hat{Y} = \bar{y}$$

TABLA 4.2. Predicción de la línea de base con el uso de la media de la variable criterio

Valor teórico de regresión: $Y = y$
Ecuación de predicción: $Y = 7$

Familia ID	*Número de tarjetas de crédito utilizadas*	*Predicción de la línea base*[a]	*Error de predicción*[b]	*Error de predicción elevado al cuadrado*
1	4	7	−3	9
2	6	7	−1	1
3	6	7	−1	1
4	7	7	0	0
5	8	7	+1	1
6	7	7	0	0
7	8	7	+1	1
8	10	7	+3	9
Total	56		0	22

[a] Número medio de tarjetas de crédito utilizadas = 56 ÷ 8 = 7.
[b] Error de predicción referido al valor real de la variable dependiente menos el valor de predicción.

Pero el investigador debe contestar todavía una cuestión: ¿qué precisión tiene la predicción? Dado que la media no dará una predicción perfecta de cada valor de la variable criterio, tenemos que crear alguna manera de valorar la exactitud de predicción, que se podría usar tanto con la predicción de la línea de base como con los modelos de regresión que creamos. El modo habitual de evaluar la adecuación de una variable predictor es examinar los errores en la predicción de la variable criterio cuando se usa para la predicción. Por ejemplo, con nuestra predicción decimos que cada familia usa siete tarjetas de crédito, de forma que estamos sobrestimando el número de tarjetas de crédito utilizado por la familia 1 en tres (véase Tabla 4.2). Por tanto, el error es +3. Si este procedimiento fuese seguido para cada familia, algunas estimaciones serían demasiado altas, otras serían demasiado bajas y a la vez otras podrían ser correctas. Aunque podríamos esperar la obtención de una medida útil de exactitud de predicción con una simple suma de los errores, esto no sería de utilidad porque los errores que proceden del uso del valor medio siempre sumarían cero. Por tanto, la suma simple de errores nunca cambiaría, independiente del grado de éxito que tuvimos con la predicción de la variable criterio con el uso de la media. Para solucionar este problema, elevamos al cuadrado el error y sumamos los resultados. El total, denominado como la **suma de los errores al cuadrado (SSE)**, proporciona una medida de la precisión predictiva que varía según la cantidad de **errores de predicción**. El objetivo es obtener la suma de los errores al cuadrado más pequeña posible, dado que esto indicaría que nuestras predicciones serían las más precisas.

Elegimos la media aritmética porque siempre producirá una suma de errores al cuadrado más pequeña que cualquier otra medida de tendencia central incluida la mediana, la moda, cualquier otro valor único o cualquier otra medida estadística más sofisticada. (Se anima a los lectores interesados a observar si pueden encontrar un mejor valor de la predicción que la media.) Por tanto, para nuestra encuesta de ocho familias, la utilización de la media como nuestra línea básica de predicción nos proporciona el mejor predictor único del número de tarjetas de crédito con una suma de errores al cuadrado de 22 (véase Tabla 4.2). En nuestra discusión de la regresión simple o múltiple, utilizaremos esta predicción a partir de la media como argumento para la comparación, dado que representa la mejor predicción sin utilizar variables independientes.

Predicción mediante una única variable independiente. Regresión simple

Como investigadores, siempre estamos interesados en mejorar nuestras predicciones. En la sección precedente, aprendimos que la media es el mejor predictor si no utilizamos otras variables inde-

pendientes. Pero en nuestra encuesta de ocho familias también recogimos información sobre otras medidas que podrían actuar como variables independientes. Determinemos si el conocimiento de una de estas variables independientes nos ayudará en nuestras predicciones por lo que se refiere a la *regresión simple*.

La regresión simple es otro procedimiento para predecir datos (al igual que la media predice datos), y utiliza la misma regla —minimizar la suma de los errores cuadrados de la predicción. Sabemos que sin utilizar el tamaño de la familia podemos describir mejor el número de tarjetas de crédito mantenidas como el valor de la media, siete. El objetivo del investigador para la regresión simple es encontrar una variable independiente que mejore la predicción de la línea de base.

El papel del coeficiente de correlación

Utilizando nuestra información sobre el tamaño de la familia, podríamos intentar mejorar nuestras predicciones reduciendo nuestros errores de predicción. Para hacerlo, los errores de predicción en el número de tarjetas de crédito mantenidas *debe* estar asociado (correlacionado) con el tamaño de la familia. El concepto de correlación, representado por el **coeficiente de correlación (*r*)**, es fundamental para el análisis de regresión y describe la relación entre dos variables. Se dice que dos variables están correlacionadas si los cambios en una variable están asociados con los cambios en la otra variable. De esta forma, a medida que una variable cambia, sabríamos cómo está cambiando la otra. Si el tamaño de la familia está correlacionado con el uso de tarjetas de crédito, escribiríamos entonces la relación como sigue:

$$\begin{array}{l}\text{Número previsto de} \\ \text{tarjetas de crédito}\end{array} = \begin{array}{l}\text{Cambio en el número de} \\ \text{tarjetas mantenidas asociadas} \\ \text{con cambio unitario en } V_i\end{array} \times \text{Valor de } V_i$$

o

$$\hat{Y} = b_1 V_i$$

En la Tabla 4.3 se muestra una ilustración del procedimiento para algunos datos hipotéticos con una única variable independiente X_1. Si encontramos que conforme aumenta X_1 en una unidad,

TABLA 4.3. Mejorando la exactitud de predicción con la adición de una constante en una ecuación de regresión

PARTE A: PREDICCIÓN SIN LA CONSTANTE
Ecuación de predicción: $Y = 2X_1$

Valor de X_1	*Variable dependiente*	*Predicción*	*Error de predicción*
1	4	2	2
2	6	4	2
3	8	6	2
4	10	8	2
5	12	10	2

PARTE B: PREDICCIÓN CON UNA CONSTANTE DE 2,0
Ecuación de predicción: $Y = 2{,}0 + 2X_1$

Valor de X_1	*Variable dependiente*	*Predicción*	*Error de predicción*
1	4	2	0
2	6	6	0
3	8	8	0
4	10	10	0
5	12	12	0

aumenta la variable criterio (sobre la media) por dos, entonces podríamos hacer predicciones para cada valor de la variable independiente. Por ejemplo, cuando X_1 tiene un valor de 4, nuestra predicción tendría un valor de 8 (véase Parte A, Tabla 4.3). Por tanto, el valor de predicción siempre es dos veces el valor de X_1 ($2X_1$). Sin embargo, muchas veces, nos encontramos que la predicción esta mejorada por la adición de un valor constante. En la Parte A de la Tabla 4.3 podemos observar que la simple predicción de dos veces X1 es errónea en cada caso. Por tanto, si cambiamos nuestra descripción para añadir una constante de dos a cada predicción, nos proporciona predicciones perfectas en todos los casos (véase Parte B, Tabla 4.3). Observaremos que cuando se estima una ecuación de regresión, normalmente merece la pena incluir una constante.

Especificación de la ecuación de regresión simple

Podemos seleccionar la «mejor» variable independiente en nuestro estudio del uso de tarjetas de crédito en base a las coeficientes de correlación dado que cuanto más alto es el coeficiente de correlación, más fuerte es la relación y por tanto más grande es la exactitud de predicción. La Tabla 4.4 contiene una matriz de correlaciones entre la variable criterio (Y) y las variables independientes (V_1,V_2 o V_3). Observando la primera columna, podemos ver que V_1, el tamaño de familia, tiene la correlación más alta con la variable criterio y por tanto es la mejor candidata para nuestra primera regresión simple. La matriz de correlación también contiene las correlaciones entre las variables independientes, aspecto muy importante en la regresión múltiple (dos o más variables independientes).

Ahora podemos estimar nuestro primer modelo de regresión simple para la muestra de ocho familias y ver cómo se ajusta la descripción a nuestros datos. El procedimiento es como sigue:

Número previsto de tarjetas de crédito mantenidas = Constante + Cambio en el número tarjetas de crédito con diferentes tamaños de familia × (Tamaño de familia)

o

$$\hat{Y} = b_0 + b_1 V_1$$

En la ecuación de regresión, representamos la constante como b_0 y la denominación b_1 se llama **coeficiente de regresión,** denotando el cambio estimado en la variable criterio por un cambio unitario de la variable independiente. El **error de predicción,** la diferencia entre los valores reales y de predicción de la variable criterio se denomina **residuo** (e). El análisis de regresión también permite que las pruebas estadísticas de la constantes y los coeficientes de regresión pueden determinar si son sustancialmente diferentes de cero (es decir, que tienen un impacto diferente al cero).

Utilizando un procedimiento matemático conocido como **mínimos cuadrados** [8, 11, 15], podemos estimar los valores de b_0 y b_1 de tal forma que la suma de los errores cuadrados de la predicción se minimiza. Para este ejemplo, los valores apropiados son una constante (b_0) de 2,87 y un coeficiente de regresión (b_1) de 0,97 por tamaño de familia. La ecuación indica que por cada miembro adicional de familia, la posesión de tarjetas de crédito es más alta como media un 0,97. Sólo

TABLA 4.4. Matriz de correlación por el estudio de uso de tarjetas de crédito

Variable	Y	V_1	V_2	V_3
Y Número de tarjetas de crédito utilizadas	1,000			
V_1 Tamaño de familia	0,866	1,000		
V_2 Renta familiar	0,829	0,673	1,000	
V_3 Número de automóviles	0,342	0,192	0,301	1,000

TABLA 4.5. Resultados de la regresión simple con el uso del tamaño de familia como la variable independiente

Valor teórico de regresión: $Y = b_0 + b_1V_1$
Ecuación de predicción: $Y = 2{,}87 + 0{,}97V_1$

Familia ID	*Número de tarjetas de crédito utilizadas*	*Tamaño de familia* (V_1)	*Predicción de regresión simple*	*Error de predicción*	*Error de predicción al cuadrado*
1	4	2	4,81	−0,81	0,66
2	6	2	4,81	1,19	1,42
3	6	4	6,75	−0,75	0,56
4	7	4	6,75	0,25	0,06
5	8	5	7,72	0,28	0,08
6	7	5	7,72	−0,72	0,52
7	8	6	8,69	−0,69	0,48
8	10	6	8,69	1,31	1,72
Total					5,50

se puede interpretar la constante 2,87 dentro de la gama de valores para la variable independiente. En este caso, un tamaño de familia de cero no es posible, por lo que la constante por sí sola no tiene un sentido práctico. Sin embargo, esto no anula su uso, dado que ayuda en la predicción de uso de tarjetas de crédito para cada tamaño de familia posible (en nuestro ejemplo de 1 a 5). En los casos en los que las variables independientes pueden adquirir valores de cero, la constante tiene una interpretación directa. Para algunas situaciones especiales donde se conoce que la relación específica pasa por el origen, la denominación de constante podría ser eliminada (denominado «regresión en el origen»). En estos casos, la interpretación de los residuos y los coeficientes de regresión cambia ligeramente. Se muestra la ecuación de regresión simple y las predicciones y residuos para cada una de las ocho familias en la Tabla 4.5.

Dado que hemos utilizado el mismo criterio (minimizar la suma de los errores al cuadrado o **mínimos cuadrados**), podemos determinar si nuestro conocimiento del tamaño familiar nos ha ayudado a predecir mejor la posesión de tarjetas de crédito cuando se compara la predicción de regresión simple con la predicción de la línea básica. La suma de los errores al cuadrado utilizando la media era 22. Ahora, la suma de los errores al cuadrado es 5,50 (véase Tabla 4.5). Utilizando el procedimiento de los mínimos cuadrados y una única variable independiente, vemos que nuestra nueva aproximación, la regresión simple, es mejor que usar sólo la media.

La creación de un intervalo de confianza para la predicción

Dado que no podemos conseguir predicciones perfectas de la variable dependiente, podríamos desear estimar el rango de valores que la variable a predecir puede tomar, en lugar de basarnos exclusivamente en una estimación simple (puntual). La estimación puntual es nuestra mejor estimación de la variable dependiente y puede demostrarse que va a ser la mejor predicción para cualquier valor dado de la variable independiente. Utilizando esta estimación puntual, podemos calcular el rango de los valores a predecir basándose en una medida de los errores de predicción que esperamos realizar. Conocido como el **error estándar de la estimación (SEE)**, esta medida es, sencillamente, la desviación estándar de los errores de predicción. Recordemos de la estadística elemental que podemos construir un intervalo de confianza para una variable sobre su valor medio añadiendo (más o menos) un cierto número de desviaciones estándar. Por ejemplo, añadiendo (más o menos) 1,96 desviaciones estándar de la media, se define un rango que incluye el 95 por ciento de los valores de una variable.

Podemos seguir un método similar para las predicciones que realizamos de un modelo de regresión. Utilizando la estimación puntual, podemos añadir (más o menos) un cierto número de errores estándar de la estimación (dependiendo del nivel de confianza deseado y del tamaño muestral) para establecer los límites superiores e inferiores de nuestras predicciones hechas con cualquier variable(s) independiente(s). El error estándar de la estimación (SEE) se calcula mediante:

$$\text{Error estándar de estimación } (SEE) = \sqrt{\frac{\text{Suma de errores al cuadrado}}{\text{Tamaño muestral} - 2}}$$

El número de SEE utilizados para derivar el intervalo de confianza se determina por el nivel de significación (α) y el tamaño muestral (N), que da un valor t. El intervalo de confianza se calcula entonces con el límite inferior siendo igual al valor previsto menos ($SEE \times$ valor t) y se calcula el límite superior como el valor previsto más ($SEE \times$ valor t). Para nuestro ejemplo de la regresión simple, $SEE = 0{,}957$ (la raíz cuadrada del valor 5,50 dividido por 6). Se construye el intervalo de confianza para las predicciones seleccionando el número de errores estándar a añadir (más/menos) mediante la búsqueda en una tabla para la distribución t y la selección del valor para una nivel de confianza dado con 6 grados de libertad (tamaño muestral menos el número de coeficientes, o $8 - 2 = 6$) es 2,447. La cantidad añadida (más/menos) al valor previsto es entonces ($0{,}957 \times 2{,}447$), o 2,34. Si sustituimos el tamaño medio de las familias (4,25) en la ecuación de regresión, entonces el valor previsto es 6,99 (difiere de la media de siete sólo en una centésima). El rango esperado va entonces de 4,65 (6,99 – 2,34) a 9,33 (6,99 + 2,34). Para una discusión más detallada de estos intervalos de confianza, véase Neter *et al.* [11].

Valoración de la exactitud de predicción

Si la suma de los errores al cuadrado (SSE) representa una medida de nuestros errores de predicción, deberíamos ser capaces de determinar una medida de nuestro éxito predictivo, que llamaremos la **suma de los cuadrados de la regresión (SSR)**. Conjuntamente, estas dos medidas deberían igual a la **suma total de los cuadrados (TSS)**, el mismo valor que nuestra predicción de la línea de base. En la medida en que el investigador añade variables independientes, el total de la suma de los cuadrados puede ahora dividirse en (1) la suma de los cuadrados prevista por la variable independiente, también conocida como la suma de los cuadrados de la regresión y (2) la suma de los errores al cuadrado:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$TSS = SSE + SSR$$

Suma total de los cuadrados	=	Suma de los errores al cuadrado	+	Suma de los cuadrados de la regresión

donde

$\bar{y}$ = media de todas las observaciones
y_i = valor de las observaciones individuales
$\hat{y}$ = valor previsto para la observación

Podemos utilizar esta división de la suma total de los cuadrados para aproximar hasta qué punto el valor teórico de la regresión describe la posesión familiar de tarjetas de crédito. El número medio de tarjetas de crédito mantenidas por nuestra muestra de familias es nuestro mejor estimador del número mantenido por cualquier familia. Sabemos que esto no es una estimación extre-

madamente precisa, pero es la mejor predicción disponible sin usar otras variables. La predicción de la línea básica con la utilización de la media fue calculada mediante el cálculo de la suma de los errores al cuadrado en la predicción (suma de los cuadrados = 22). Ahora que hemos ajustado un modelo de regresión utilizando el tamaño de la familia, ¿se explica así la variación mejor que con la media? Sabemos que es algo mejor porque la suma de los errores al cuadrado es ahora 5,50. Podemos observar así el alcance de nuestro modelo con la investigación de esta mejora.

Suma de los errores al cuadrado (predicción de la línea básica)	(SS_{Total} o SST)	22,0
– Suma de los errores al cuadrado (regresión simple)	(SS_{Error} o SSE)	–5,5
Suma de los errores al cuadrado explicados (regresión simple)	($SS_{\text{Regresión}}$ o SSR)	16,5

Por tanto, explicamos 16,5 errores al cuadrado cambiando de la media a un modelo de regresión utilizando el tamaño de la familia. Esto supone una mejora del 75 por ciento (16,5 ÷ 22 = 0,75) sobre el uso de la línea básica. Otra forma de expresar este nivel de precisión predictiva es el **coeficiente de determinación (R^2)**, el ratio de la suma al cuadrado de la regresión sobre el total de la suma de los cuadrados como muestra la siguiente ecuación:

$$\text{Coeficiente de determinación } (R^2) = \frac{\text{Suma de los cuadrados de la regresión}}{\text{Suma del total de los cuadrados}}$$

Teniendo en cuenta que el modelo de regresión que utiliza el tamaño de la familia predice perfectamente todas las tarjetas de crédito mantenidas por las familias, $R^2 = 1{,}0$ y que la utilización del tamaño de la familia no ofreció mejores predicciones que utilizando la media, $R^2 = 0$, cuando la ecuación de la regresión contiene más de una variable independiente, el valor del R^2 representa el efecto combinado del valor teórico en el conjunto en la predicción. El valor del R^2 es simplemente la correlación al cuadrado de los valores reales y los valores previstos.

Cuando se utiliza el coeficiente de correlación (r) para evaluar la relación entre las variables dependientes e independientes, el signo del coeficiente de correlación ($+r$, $-r$) denota la pendiente de la línea de regresión. Sin embargo, la «fuerza» de la relación se representa mejor por el R^2, que es, por supuesto, siempre positiva. En nuestro ejemplo el $R^2 = 0{,}75$, que indica que el 75 por ciento de la variación en la variable dependiente se explica por la variable independiente. Cuando las discusiones mencionan la variación de la variable dependiente, se refieren a esta suma total de cuadrados que el análisis de regresión intenta predecir con una o más variables independientes.

Predicción utilizando varias variables independientes: Análisis de regresión múltiple

Hemos demostrado previamente cómo una regresión simple mejora nuestra predicción de la posesión de tarjetas de crédito. Usando los datos del tamaño de la familia, predecimos el número de tarjetas de crédito que poseería una familia mejor de lo que podríamos realizar utilizando simplemente la media aritmética. Este resultado pone de manifiesto la cuestión de si pudiésemos mejorar nuestra predicción utilizando datos adicionales obtenidos de las mismas familias. ¿Mejoraría nuestra predicción del número de tarjetas de crédito si utilizáramos datos del tamaño de la familia además de datos de otra variable, quizá la renta familiar?

El impacto de la multicolinealidad

La capacidad de una variable independiente adicional de mejorar la predicción de una variable criterio tiene relación no sólo con la correlación con la variable dependiente, sino también respecto de las correlaciones de las variables independientes adicionales en función de las variables inde-

pendientes ya presentes en la ecuación de regresión. **Colinealidad** es la asociación, medida como correlación, entre dos variables independientes. **Multicolinealidad** se refiere a la correlación entre tres o más variables independientes (evidenciada cuando se hace la regresión de una respecto de las otras). Aunque existe una distinción precisa en términos estadísticos, es muy común en la práctica utilizar estos términos indistintamente.

El impacto de la multicolinealidad consiste en reducir el poder predictivo de cualquier variable independiente individual en la medida en que está asociado con las otras variables independientes. Conforme se incrementa la colinealidad, la varianza única explicada por cada variable independiente se reduce y el porcentaje de predicción compartida aumenta. Dado que sólo se puede calcular una vez se ha realizado la predicción compartida, la predicción global aumenta mucho más lentamente conforme se añaden variables independientes con un nivel de multicolinealidad alta. Para maximizar la predicción de un número específico de variables independientes, el investigador debería buscar otras variables independientes que tienen una multicolinealidad baja con las otras variables independientes pero que también tienen correlaciones altas con la variable dependiente. Trataremos de nuevo las cuestiones de colinealidad y multicolinealidad en secciones posteriores para abordar sus implicaciones para la selección de variables independientes y la interpretación del valor teórico de regresión.

La ecuación del análisis de regresión múltiple

Para mejorar aún más nuestra predicción de la posesión de tarjetas de crédito, utilicemos unos datos adicionales obtenidos de nuestras ocho familias. La segunda variable independiente para ser incluida en el modelo de regresión es la renta familiar (V_2), que tiene la siguiente correlación más alta con la variable dependiente. Aunque V_2 tiene bastante correlación con V_1 (que ya se contempla en la ecuación), es todavía la siguiente mejor variable para entrar porque V_3 tiene una correlación mucho más baja con la variable dependiente. Simplemente, ampliamos nuestro modelo de regresión simple para incluir dos variables independientes como sigue:

$$\text{Predicción del número de tarjetas de crédito utilizadas} = b_0 + b_1V_1 + b_2V_2 + e$$

donde

b_0 = número constante de tarjetas de crédito independientemente del tamaño y renta familiar
b_1 = cambio en la posesión de tarjetas de crédito asociado a un cambio unitario en el tamaño familiar
b_2 = cambio en la posesión de tarjetas de crédito asociado con un cambio unitario en la renta familiar
V_1 = tamaño de la familia
V_2 = renta de la familia

El modelo de regresión múltiple con dos variables independientes, cuando se estima con el procedimiento de mínimos cuadrados, tiene un constante de 0,482 con unos coeficientes de regresión de 0,63 y 0,216 para V_1 y V_2 respectivamente. De nuevo podemos hallar nuestro residuo en la predicción de Y, restando de nuestra predicción el valor efectivo. Elevamos entonces al cuadrado el error de predicción resultante, como en la Tabla 4.6. La suma de los errores al cuadrado es de 3,04 para nuestra predicción utilizando tanto la renta familiar como el tamaño de la familia. Se puede comparar con el valor del modelo de regresión simple de 5,50 (Tabla 4.5) utilizando sólo el tamaño familiar para la predicción.

Cuando se añade al análisis de la regresión la renta de la familiar, el R^2 aumenta a 0,86:

$$R^2_{\text{(tamaño de la familia + renta familiar)}} = \frac{22{,}0 - 3{,}04}{22{,}0} = \frac{18{,}96}{22{,}0} = 0{,}86$$

TABLA 4.6. Resultados de la regresión múltiple utilizando el tamaño familiar y la renta familiar como variables independientes

Valor teórico de regresión: $Y = b_0 + b_1V_1 + b_2V_2$
Ecuación de predicción: $Y = 0{,}482 + 0{,}63V_1 + 0{,}216V_2$

Familia ID	*Número de tarjetas de crédito utilizadas*	*Tamaño de familia (V_1)*	*Renta familiar (V_2)*	*Predicción de regresión simple*	*Error de predicción*	*Error de predicción al cuadrado*
1	4	2	14	4,76	–0,76	0,58
2	6	2	16	5,20	0,80	0,64
3	6	4	14	6,03	–0,03	0,00
4	7	4	17	6,68	0,32	0,10
5	8	5	18	7,53	0,47	0,22
6	7	5	21	8,18	–1,18	1,39
7	8	6	17	7,95	0,05	0,00
8	10	6	25	9,67	0,33	0,11
Total						3,04

Esto significa que la inclusión de la renta familiar en el análisis de regresión aumenta la predicción en un 11 por ciento (0,86 – 0,75), debido al incremento de la potencia predictiva de la renta familiar.

La adición de una tercera variable independiente

Hemos observado un incremento en la exactitud de predicción que se gana con el cambio de la ecuación de regresión simple a la ecuación de regresión múltiple, pero también tenemos que tener en cuenta que en algún momento la adición de variables independientes también será menos ventajosa e incluso en algunos casos es contraproducente. En nuestra encuesta de la utilización de tarjetas de crédito, tenemos otra adición posible de la ecuación de regresión múltiple, el número de posesión de automóviles (V_3). Si ahora especificamos la ecuación de regresión para incluir todas las tres variables independientes, podemos observar alguna mejora en la ecuación de regresión, pero no de la misma envergadura vista anteriormente. El valor R^2 aumenta a 0,87, lo que representa sólo un incremento del 0,01 sobre el modelo anterior de regresión múltiple. Además, tal y como se aborda en una sección posterior, el coeficiente de regresión para el V_3 no es estadísticamente significativo. Por tanto, en este caso, el investigador hará mejor en emplear el modelo de regresión múltiple con dos variables independientes (tamaño de familia y renta) y no utilizar la tercera variable independiente (número de posesión de automóviles) para hacer predicciones.

Resumen

El análisis de regresión es una técnica de dependencia simple y sencilla que puede proporcionar al investigador tanto predicción como explicación. El ejemplo previo ha ilustrado los conceptos y procedimientos básicos del análisis de regresión con el fin de desarrollar un conocimiento de la racional y las cuestiones de este procedimiento en su forma más básica. Las siguientes secciones tratan estas cuestiones en más detalle y proporcionan un proceso de decisión para aplicar el análisis de regresión a cualquier problema de investigación.

Un proceso de decisión para el análisis de regresión múltiple

En las secciones previas discutimos ejemplos de regresiones simples y múltiples. En aquellas discusiones, diversos factores afectaban a nuestra capacidad para encontrar el mejor modelo de re-

gresión. Hasta este punto, sin embargo, hemos examinado estos supuestos sólo en términos sencillos, con escasa relación a cómo se combinan en una aproximación global al análisis de la regresión múltiple. En las siguientes secciones, utilizaremos el proceso de modelización en seis pasos que introdujimos en el Capítulo 1 como esquema para discutir los factores que afectan a la creación, estimación, interpretación y validación de un análisis de regresión. El proceso comienza con la especificación de los objetivos del análisis de regresión, incluyendo la selección de las variables dependientes e independientes. Entonces el investigador procede a diseñar el análisis de regresión, con la consideración de factores tales como el tamaño muestral y la necesidad de transformaciones de variables. Una vez formulado el modelo de regresión, se contrastan en primer lugar los supuestos subyacentes al análisis de regresión para las variables individualmente. Si se cumplen todos los supuestos, entonces es cuando se estima el modelo.Una vez que se obtienen los resultados, se lleva a cabo análisis de diagnostico para asegurar que el modelo global cumple los supuestos de regresión y que ninguna observación tiene una influencia indebida sobre los resultados. El siguiente paso es la interpretación del valor teórico de la regresión, donde se examina el papel jugado por cada variable independiente en la predicción de la medida dependiente. Finalmente, los resultados se validan para asegurar la generalidad a su población. La Figura 4.1 y 4.6 proporcionan los pasos 1-3 y 4-6 respectivamente de una representación gráfica de proceso de construcción del modelo de regresión múltiple. En las siguientes secciones se discutirá cada paso con detalle.

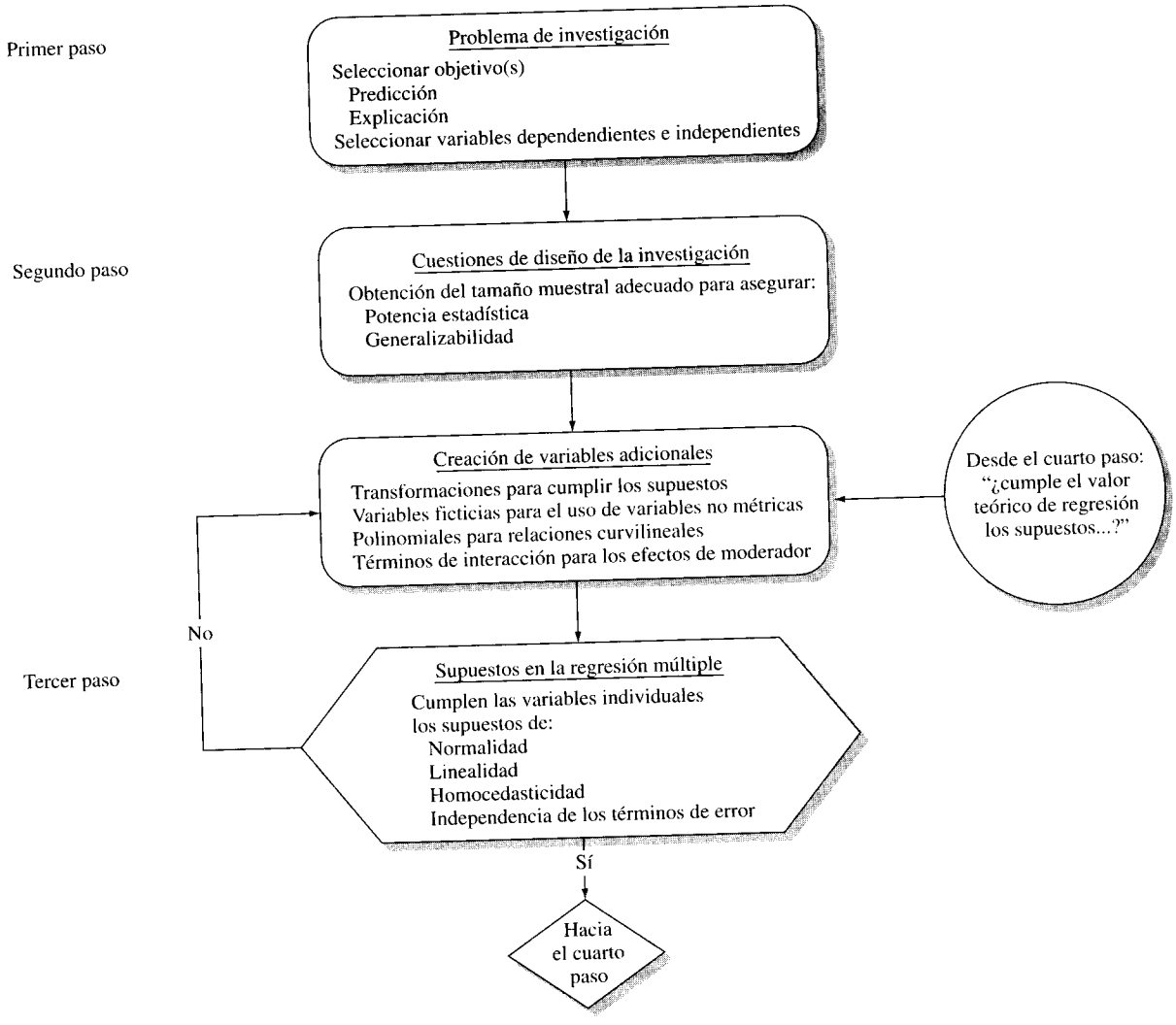


FIGURA 4.1. Pasos 1-3 en el diagrama de decisión de la regresión múltiple.

Primer paso: Objetivos de la regresión múltiple

El análisis de regresión múltiple, una forma de modelo lineal general, es una técnica estadística de análisis multivariante utilizada para examinar las relaciones entre una única variable criterio y un conjunto de variables independientes. El punto de partida necesario del análisis multivariante, como en todas las técnicas estadísticas multivariantes, es el problema a investigar. La flexibilidad y la capacidad de adaptación de la regresión múltiple permite utilizarlos con casi cualquier relación de dependencia. En la selección de la aplicación adecuada del análisis de regresión, el investigador debe considerar tres asuntos fundamentales: (1) la conveniencia del programa de investigación, (2) la especificación de una relación estadística y (3) la selección de las variables dependientes e independientes.

Problemas de investigación adecuados para la regresión múltiple

La regresión múltiple es con mucho la técnica univariante más utilizada de todas las examinadas en este texto. Con sus amplias posibilidades de aplicación, el análisis de regresión múltiple se utiliza para diversos propósitos. Las crecientes aplicaciones de la regresión múltiple, sin embargo, se agrupan en dos amplias clases de problemas de investigación: predicción y explicación. Estos problemas de investigación no son mutuamente excluyentes, y se puede llevar a cabo una aplicación del análisis de regresión múltiple para cualquiera de los dos tipos de problemas de investigación.

Predicción con regresión múltiple

Un propósito fundamental de una regresión múltiple es la predicción de la variable criterio con un conjunto de variables independientes. Al hacerlo, la regresión múltiple cumple uno de dos objetivos. El primer objetivo es maximizar la potencia conjunta de predicción de las variables independientes tal y como se representan en el valor teórico. Esta combinación lineal de variables independientes se construye de tal forma que se convierta en un predictor óptimo de la variable criterio. La regresión múltiple proporciona un medio objetivo de evaluar el poder predictivo de un conjunto de variables independientes. En aplicaciones centradas en este objetivo, el investigador está principalmente interesado en conseguir la máxima predicción. La regresión múltiple proporciona muchas opciones tanto en la forma como en la especificación de las variables independientes que puedan modificar el valor teórico para aumentar su poder predictivo. Muchas veces la predicción se maximiza a expensas de la interpretación. Un ejemplo es una variante del análisis de regresión, el análisis de series temporales, en el cual el único propósito es predecir y la interpretación de los resultados es útil sólo como un medio de incrementar la precisión predictiva. En otras situaciones, la precisión predictiva es crucial para asegurar la validez del conjunto de variables independientes, teniendo en cuenta la ulterior interpretación del valor teórico. Las medidas de precisión predictiva y los test estadísticos se forman en relación con la significación del poder predictivo que pueda obtenerse. En todos los casos, tanto si la predicción es o no es el objetivo principal, el análisis de regresión debe conseguir niveles aceptables de precisión predictiva para justificar su aplicación. El investigador debe asegurarse de tener en cuenta tanto la significación práctica como la estadística (véase el paso cuarto, a continuación de esta discusión).

La regresión múltiple puede también conseguir un segundo objetivo de comparación de dos o más conjuntos de variables independientes para averiguar el poder predictivo de cada valor teórico. Ilustrativo de una aproximación de modelización confirmatoria, este uso de la regresión múltiple se centra en la comparación de resultados entre dos o más alternativas o modelos en competencia. El objetivo principal de este tipo de análisis es el poder predictivo relativo entre modelos, aunque en cualquier situación la predicción del modelo elegido debe demostrar tanto significación práctica como estadística.

Explicación con regresión múltiple

La regresión múltiple proporciona también un medio de evaluar objetivamente el grado y carácter de la relación entre las variables dependientes e independientes al formar el valor teórico. Las variables independientes, además de su predicción conjunta de la variable dependiente, pueden considerarse también por su contribución individual al valor teórico y a sus predicciones. La interpretación del valor teórico puede tomarse desde alguna de estas tres perspectivas: la importancia de las variables independientes, los tipos de relaciones encontradas o las interrelaciones entre las variables independientes.

La interpretación más directa del valor teórico de la regresión es una determinación de la importancia relativa de cada variable independiente en la predicción de la medida independiente. En todas las aplicaciones, la selección de variables independientes se basaría en sus relaciones teóricas con la variable dependiente. El análisis de regresión proporciona un medio de evaluar objetivamente la magnitud y dirección (positiva o negativa) de cada relación con la variable independiente. El carácter de la regresión múltiple, que la diferencia de sus contrapartidas univariantes, es la evaluación simultánea de relaciones entre cada variable independiente y las medidas de la dependiente. Al realizar esta evaluación simultánea, se determina la importancia relativa de cada predictor.

Además de evaluar la importancia de cada variable, la regresión múltiple permite también al investigador la evaluación de la naturaleza de las relaciones entre las variables independientes y la variable dependiente. La relación supuesta es una asociación lineal basada en correlaciones entre las variables independientes y la variable dependiente. Pero también se disponen de transformaciones para evaluar si existen otros tipos de relación, particularmente las relaciones curvilineales. Esta flexibilidad asegura que el investigador pueda examinar la verdadera naturaleza de las relaciones más allá de la supuesta relación lineal.

Finalmente, la regresión múltiple proporciona también una idea de las relaciones entre las variables independientes en sus predicciones de la variable dependiente. Estas interpretaciones son importantes por dos razones. En primer lugar, la correlación entre las variables independientes puede hacer que algunas variables sean redundantes en su esfuerzo predictivo. Como tal, no son necesarias para producir una predicción óptima. No se trata de reflejar sus relaciones individuales con la variable criterio sino que indica que en un contexto multivariante, no son necesarias si se emplea otro conjunto de variables independientes para explicar esta varianza. El investigador debe mostrarse precavido con la determinación de la importancia de las variables independientes basadas solamente en el valor teórico derivado, porque las relaciones entre las variables independientes pueden «enmascarar» relaciones que no se necesitan para propósitos predictivos pero que representan sin embargo hallazgos sustantivos. Las interrelaciones entre las variables pueden extenderse no sólo a su poder predictivo sino también a las interrelaciones entre sus efectos estimados. Esto se ve mejor cuando el efecto de una variable independiente es contingente con otra variable independiente. La regresión múltiple proporciona un diagnóstico que puede determinar si existen tales efectos basados en razones empíricas o teóricas. Las indicaciones de un alto grado de interrelaciones (multicolinealidad) entre las variables independientes pueden sugerir el uso de las escalas sumadas, tal y como se abordó en el Capítulo 2.

Especificación de la relación estadística

Una regresión múltiple es apropiada cuando el investigador está interesado en una relación estadística, no funcional, Por ejemplo, examinemos la siguiente relación:

Coste total = Coste variable + Coste fijo

Si el coste variable es de 2$ por unidad, el coste fijo es de 500$ y producimos 100 unidades, suponemos que el coste total será exactamente 700$ y que cualquier desviación de los 700$ está causada por nuestra incapacidad para medir el coste dado que la relación entre costes es fija. A esto

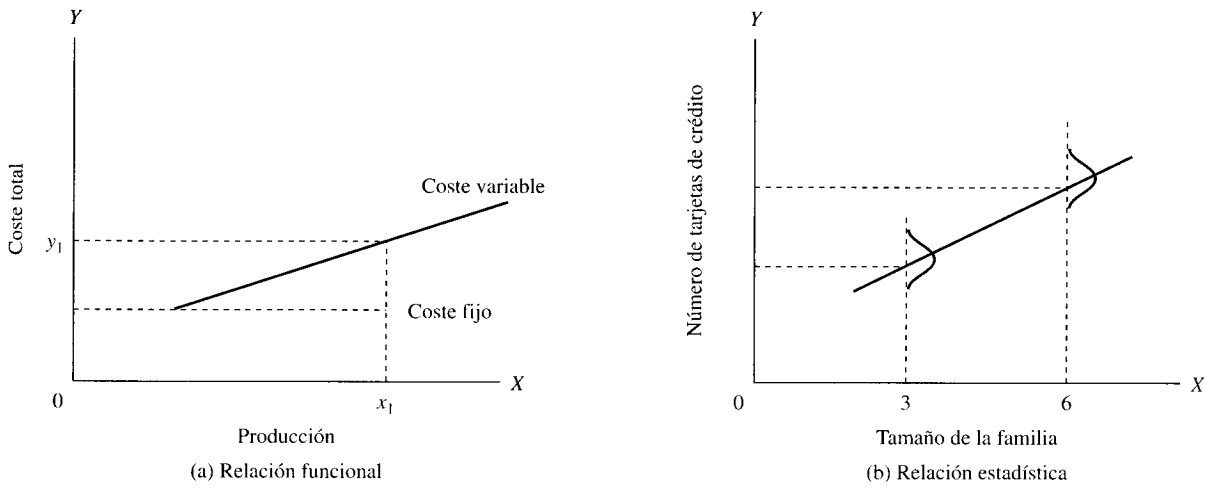


FIGURA 4.2. Comparación de las relaciones funcionales y estadísticas.

se le denomina una *relación funcional* porque esperamos que no existirá un error en nuestra predicción.

Pero en nuestro ejemplo anterior sobre la muestra de datos que representa el comportamiento humano, estábamos suponiendo que nuestra descripción del uso de las tarjetas de crédito era sólo aproximada y no una predicción perfecta. Se pensaba que era una **relación estadística** porque siempre existiría un componente aleatorio en la relación examinada. Encontramos dos familias con dos miembros, dos con cuatro miembros, etc., que tenían distinto número de tarjetas de crédito. En una relación estadística se observará más de un valor de la variable dependiente para cualquier valor de una variable predictor. La variable criterio se supone que es una variable aleatoria, y para un predictor dado sólo podemos esperar estimar el valor medio de la variable criterio asociado con él. En nuestro ejemplo de la regresión simple, las dos familias con cuatro miembros mantienen una media de 6,5 tarjetas de crédito, y nuestra predicción era de 6,75. Nuestra predicción no es tan precisa como desearíamos, pero es mejor que usar nada más que la media de 7 tarjetas de crédito. Se supone que el error es el resultado de un comportamiento aleatorio entre los poseedores de las tarjetas.

En resumen, una relación funcional calcula un valor exacto, mientras que una relación estadística estima un valor medio. A lo largo del libro, nos centraremos en las relaciones estadísticas. Los dos tipos de relación se representan en la Figura 4.2.

Selección de variables dependientes e independientes

El «éxito» final de cualquier técnica multivariante, incluyendo las regresiones múltiples, comienza con la selección de las variables que se van a usar en el análisis. Dado que la regresión múltiple muestra una relación de dependencia, el investigador debe especificar qué variable es la que se usa como criterio y qué variables se usan como predictor. La selección de ambos tipos de variables debería basarse principalmente en fundamentos conceptuales o teóricos. Los Capítulos 1 y 11 exponen el papel de la teoría en el análisis multivariante, y todos aquellos asuntos fuertemente relacionados con la regresión múltiple. El investigador debe tomar las decisiones fundamentales de la selección de variables, incluso aunque tenga muchas opciones y comandos de programas para ayudarle en la estimación del modelo. Si no emite juicios durante la selección de la variable y en su lugar (1) selecciona las variables indiscriminadamente o (2) permite que la selección de una variable independiente se base exclusivamente en bases empíricas, se incumplirán varios de los principios básicos del desarrollo del modelo.

La selección de una variable criterio está muchas veces dictada por el problema de la investigación. Pero en muchos casos, el investigador debe ser consciente del **error de medida**, especialmente en la variable independiente. El error de medida se refiere al grado en que la variable es

una medida precisa y consistente del concepto que está siendo estudiado. Si la variable que se utiliza como dependiente tiene un error sustancial de medida, entonces incluso las mejores variables independientes pueden ser incapaces de conseguir niveles aceptables de precisión predictiva. El error de medida puede venir de diversas fuentes (véase Capítulo 1 para una discusión más detallada). El error de medida que es problemático puede ser abordado mediante el uso de las escalas aditivas, tal y como se comenta en los Capítulos 1 y 3. El investigador debe siempre interesarse por la obtención de la mejor medida de las variables dependientes e independientes, basadas ambas en factores empíricos y conceptuales.

El supuesto más problemático en la selección de la variable independiente es el **error de especificación**, que hace referencia a la inclusión de variables irrelevantes o a la omisión de variables relevantes del conjunto de variables independientes. Aunque la inclusión de una variable irrelevante no sesgue los resultados de las otras variables independientes, tiene cierto impacto sobre ellos. En primer lugar, reduce la parsimonia del modelo, que puede ser crítica en la interpretación de los resultados. En segundo lugar, las variables adicionales pueden enmascarar o desplazar los efectos de variables más útiles, especialmente si se utiliza alguna forma jerárquica de estimación del modelo (véase paso 4 para más detalle). Finalmente, las variables adicionales pueden hacer que las contrastaciones de la significación estadística de las variables independientes sean menos precisas y reduzcan la significación estadística y práctica del análisis.

Dados los problemas asociados con la adición de variables irrelevantes, ¿debe fijarse el investigador en las variables relevantes excluidas? La respuesta es definitivamente sí, porque la exclusión de las variables relevantes puede sesgar seriamente los resultados y afectar negativamente cualquier interpretación de ellos. En el caso más simple, las variables omitidas no están correlacionadas con las variables incluidas, y el único efecto es reducir la precisión predictiva conjunta del análisis. Pero cuando existe correlación entre las variables incluidas y las omitidas, los efectos de las variables incluidas pueden verse segados en la medida en que están correlacionadas con las variables omitidas. Cuanto mayor sea la correlación, mayor será el sesgo. Los efectos estimados para las variables incluidas representan ahora no sólo sus efectos reales sino también los efectos que las variables incluidas comparten con las variables omitidas. Esto nos puede llevar a serios problemas en la interpretación de los modelos y en la evaluación de la significación estadística y práctica.

El investigador debe ser cuidadoso en la selección de las variables para evitar ambos tipos de errores de especificación. Quizá los mayores problemas consistan en la omisión de las variables relevantes, dado que los efectos de las variables no pueden evaluarse sin su inclusión. Esto intensifica la necesidad de un soporte práctico y teórico de todas las variables incluidas o excluidas en un análisis de regresión múltiple.

Los errores de medida afectan también a las variables independientes reduciendo su poder predictivo en la medida en que aumenta el error de medida. La regresión múltiple no tiene medios directos para corregir los niveles conocidos de error de medida para las variables independientes. Si el investigador sospecha que el error de medida puede ser problemático en las variables independientes, deberían utilizarse los modelos de ecuaciones estructurales (Capítulo 11) como un medio de tratar los errores de medida en la estimación de los efectos de las variables independientes.

Segundo paso: Diseño de la investigación en el análisis de la regresión múltiple

En el diseño de un análisis de regresión múltiple, el investigador debe considerar asuntos tales como el tamaño muestral, la naturaleza de las variables independientes y la posible creación de nuevas variables para representar las especiales relaciones entre las variables dependientes e independientes. Al hacerlo, debe mantener siempre el criterio de significación práctica y estadística. La capacidad de las regresiones múltiples para realizar muchos tipos de investigaciones se ve enormemente influenciada por los supuestos del diseño de la investigación que se discutirán a continuación.

Tamaño muestral

El tamaño muestral utilizado en la regresión múltiple es quizá el elemento aislado más influyente bajo control del investigador en el diseño del análisis. Los efectos del tamaño muestral se ven más directamente en la potencia estadística del test de significación y la generalización del resultado. Trataremos ambos asuntos en las secciones siguientes.

Potencia estadística y tamaño muestral

El tamaño muestral tiene un impacto directo en la conveniencia y la potencia estadística de la regresión múltiple. Muestras pequeñas, habitualmente caracterizadas por tener menos de 20 observaciones, son apropiadas sólo para análisis de regresión simple con una única variable independiente. Incluso en estas situaciones, sólo se pueden detectar relaciones muy fuertes con cierto grado de certidumbre. De la misma forma, las muestras muy grandes, de 1.000 observaciones o más, hacen los test de significación estadística demasiado sensibles, indicando que casi cualquier relación es estadísticamente significativa.

Con muestras muy grandes el investigador debe asegurar que el criterio de significación práctica se cumpla a la vez que la significación estadística.

La **potencia** de la regresión múltiple se refiere a la probabilidad de detectar como estadísticamente significativo un nivel específico de R^2 o un coeficiente de regresión para un nivel de significación especificado y un tamaño de muestra específico (véase Capítulo 1 para una discusión más detallada). El tamaño muestral tiene un impacto directo y cuantificable sobre la potencia. La Tabla 4.7 ilustra la interacción entre el tamaño muestral, el nivel de significación (α) elegido y el número de variables independientes para detectar un R^2 significativo. Los valores de la tabla son el mínimo R^2 que el tamaño muestral especificado detectará como estadísticamente significativo y el nivel alfa especificado con una probabilidad (potencia) de 0,80. Por ejemplo, si el investigador emplea cinco variables independientes, especifica el nivel de significación de 0,05 y está satisfecho al detectar el R^2 del 80 por ciento de las veces que ocurre (correspondiente a una potencia de 0,80), una muestra de 50 encuestados detectará valores de R^2 del 23 por ciento y superior. Si la muestra aumenta en 100 encuestados, entonces se detectarán valores del R^2 del 12 por ciento o superiores. Pero si los 50 encuestados es todo lo que tiene el investigador y quiere un nivel de significación del 0,01, el investigador detectará valores de R^2 sólo por encima del 29 por ciento. El investigador debería considerar el papel del tamaño muestral en la contrastación de la significación antes de la recogida de datos. Si se esperan relaciones débiles, puede hacer juicios contrastados en la medida en que el tamaño de muestra necesario detecte razonablemente las relaciones, si existen. Por ejemplo, la Tabla 4.7 demuestra que los tamaños de la muestra de 100 detectarán claramente valores de R^2 bajos (del 10 al 15 por ciento) por encima de 10 variables independientes y un nivel de

TABLA 4.7. Mínimo R^2 que se puede encontrar estadísticamente significativo con una potencia de 0,80 para diferentes variables independientes y tamaños muestrales

	Nivel de significación (α) = 0,01 *Número de variables independientes*				*Nivel de significación (α) = 0,05* *Número de variables independientes*			
Tamaño muestral	2	5	10	20	2	5	10	20
20	45	56	71	NA	39	48	64	NA
50	23	29	36	49	19	23	29	42
100	13	16	20	26	10	12	15	21
250	5	7	8	11	4	5	6	8
500	3	3	4	6	3	4	5	9
1.000	1	2	2	3	1	1	2	2

NA = no aplicable.

significación de 0,05. Sin embargo, si el tamaño muestral es menor de 50 observaciones en estas situaciones, puede detectarse el doble de R^2. El investigador siempre debe ser consciente de la potencia anticipada de cualquier análisis de regresión propuesto y entender los elementos del diseño de la investigación que pueden cambiarse para cumplir los requerimientos de un análisis aceptable [9].

El investigador puede determinar también el tamaño muestral necesario para detectar los efectos de las variables independientes individuales dado el efecto del tamaño esperado (correlación), el nivel alfa y la potencia deseada. Los cálculos posibles son muy numerosos para presentarse en esta discusión, pero se remite al lector interesado a varios textos que analizan los análisis de potencia [5] o de un programa de computador para calcular el tamaño muestral para una situación dada [3].

Generalización y tamaño muestral

Además de tener un papel importante en la determinación de la potencia estadística, el tamaño muestral también afecta a la generalización de los resultados, en función del ratio de observaciones sobre las variables independientes. Una norma general es que este ratio nunca debería caer por debajo de cinco, lo que significa que existirán cinco observaciones para cada variable independiente presente en el valor teórico. Conforme cae este ratio por debajo de cinco, el investigador pueden incurrir en el riesgo de «sobreajustar» el valor teórico respecto de la muestra, haciendo que los resultados sean muy específicos para esa muestra y por tanto con falta de generalización. Mientras que el ratio mínimo es 5 a 1, el nivel deseado está entre 15 y 20 observaciones para cada variable independiente. Cuando este nivel se alcance, los resultados deberían ser generalizables si el tamaño muestral es representativo. Sin embargo, si se emplea un proceso por pasos (discutido en el cuarto paso bajo la aproximación a la estimación del modelo), el nivel recomendado aumenta de 50 a 1. En casos donde la muestra disponible no cumpla estos criterios, el investigador debería asegurarse la validación de la generalización de los resultados.

Predictores de efectos fijos frente a predictores de efectos aleatorios

Los ejemplos de los modelos de regresión que hemos discutido hasta este punto han sido formulados bajo el supuesto de que los niveles de las variables predictor son fijos. Por ejemplo, si queremos saber el impacto sobre los tres niveles de preferencia de un edulcorante de una bebida de cola, presentaremos tres tandas de este tipo de bebidas a un grupo de personas. Entonces predecimos el ratio de preferencia sobre cada bebida, utilizando el nivel de edulcorante como predictor. Hemos fijado el nivel de edulcorante y estamos interesados en su efecto en estos niveles. No suponemos que los tres niveles sean una muestra aleatoria de un gran número de posibles niveles de edulcorante. Una variable predictor aleatoria es aquella en que los niveles de la variable predictor se seleccionan aleatoriamente. Cuando utilizamos una variable predictor aleatoria, el interés no está sólo en los niveles examinados sino en la mayor cantidad de posibles niveles del predictor de los que se seleccionan en una muestra.

La mayor parte de los modelos de regresión basados en una encuesta de datos son modelos de efectos aleatorios. A modo de ilustración, se dirigió una encuesta para ayudar a evaluar la relación entre la edad del encuestado y la frecuencia de sus visitas al médico. La variable predictor «edad del encuestado» se seleccionó aleatoriamente de la población, y la inferencia en relación con la población es el objetivo, no sólo el conocimiento de los individuos de la muestra.

Los procedimientos de estimación de modelos que utilizan ambos tipos de variables son las mismas excepto para los términos de error. En los modelos de efectos aleatorios, una parte del error aleatorio procede del muestreo de los predictores. Sin embargo, los procedimientos estadísticos basados en el modelo fijo son bastante robustos, así que utilizar el análisis estadístico como si se estuviese tratando con un modelo fijo (como suponen la mayor parte de los programas de análisis) puede considerarse como una aproximación razonable.

Creación de variables adicionales

La relación básica representada en la regresión múltiple es la asociación *lineal* entre variables dependientes e independientes *métricas* basada en la correlación momento-producto. Muchas veces los investigadores se enfrentan al problema de incorporar datos no métricos, tales como género o ocupación, en una ecuación de regresión. Sin embargo, tal y como hemos visto anteriormente, la regresión se limita a los datos métricos. Además, la incapacidad de la regresión de modelizar directamente las relaciones no lineales puede suponer una restricción para el investigador cuando se enfrenta con situaciones en las que una relación no lineal (por ejemplo, en forma de U) es sugerida por la teoría o es detectada cuando examina los datos.

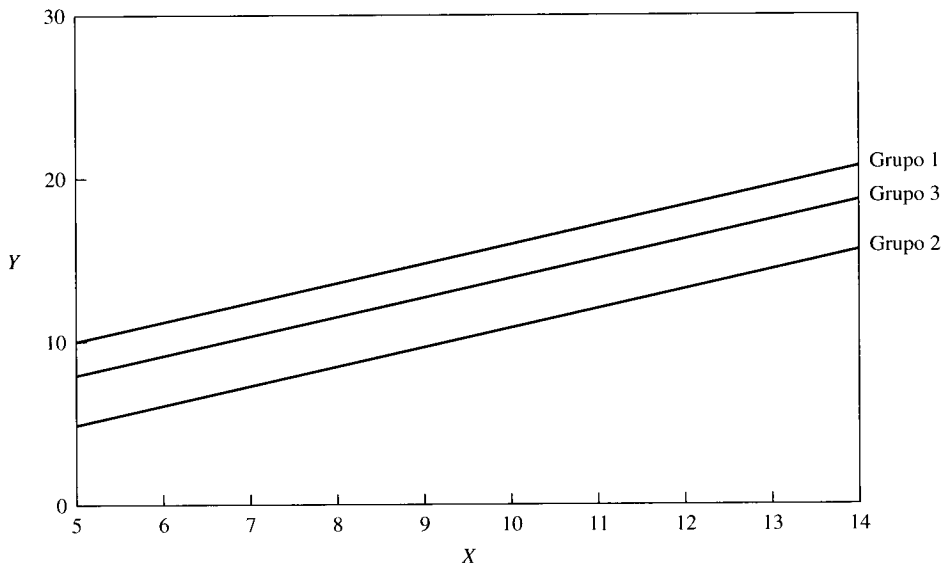
En estas situaciones, deben crearse nuevas variables mediante **transformaciones**, en la medida en que la regresión múltiple descansa completamente en los tipos de variables del modelo que sólo representan relaciones lineales. La transformación de los datos ofrece un medio de modificar tanto las variables dependiente como las independientes por una de estas dos razones: mejorar o modificar la relación entre las variables dependientes o independientes o permitir el uso de variables no métricas en el valor teórico de la regresión. Las transformaciones de los datos pueden basarse en razones tanto «teóricas» (transformaciones cuya conveniencia es sugerida estrictamente por la naturaleza de los datos) o «derivadas de los datos» (transformaciones sugeridas estrictamente por el examen de los datos). En cualquier caso el investigador debe proceder muchas veces por ensayo y error, evaluando constantemente las mejoras frente a la necesidad de transformaciones adicionales. Exploraremos estos supuestos con discusiones acerca de las transformaciones de los datos que permitan al análisis de regresión representar de la mejor forma posible los datos reales y discutir la creación de variables que complementen a las variables originales.

Todas las transformaciones descritas se pueden realizar fácilmente mediante las funciones de los programas de estadística más extendidos. Nos centramos en transformaciones que puedan calcularse de esta forma, aunque existen otros métodos de transformación de los datos más sofisticados y complicados (por ejemplo, véase Box y Cox [4]).

Incorporación de datos no métricos con variables ficticias

Una situación común a la que se enfrentan los investigadores es la presencia de variables independientes no métricas. No obstante, hasta ahora todas nuestras explicaciones han supuesto medidas métricas tanto para las variables independientes como las variables dependientes. Cuando la variable criterio se mide como una variable dicotómica (0, 1), lo más apropiado es el análisis discriminante o una forma especializada de regresión (regresión logística), ambos presentados en el Capítulo 5. Pero, ¿qué podemos hacer cuando las variables independientes son no métricas, con dos o más categorías? En el Capítulo 2 se introdujo el concepto de las variables dicotómicas, conocidas como **variables ficticias**, que actúan en lugar de las variables independientes. Cada variable representa una categoría de variable independiente no métrica y cualquier variable no-métrica con k categorías puede representarse como una variable ficticia $k - 1$.

Existen dos formas de codificación de variable ficticia. La más común es la **codificación de indicador** en la que se representa la categoría por 1 o 0. Los coeficientes de regresión para la variable ficticia representan desviaciones para cada grupo de encuestados formado por una variable ficticia de la **categoría de referencia** (es decir, el grupo omitido que recibe todos los ceros) respecto la variable dependiente. Estas diferencias de grupo pueden ser valoradas directamente, dado que los coeficientes están en las mismas unidades que la variable dependiente. Esta forma de codificación de la variable ficticia puede ser mostrada como atajos diferentes de los grupos (véase Figura 4.3). En este ejemplo, se representa una variable no métrica de tres categorías por dos variables ficticas (D_1 y D_2) dando valores a los grupos 1 y 2, respecto del grupo 3, la **categoría de referencia**. Los coeficientes de regresión son 2,0 para D_1 y –3,0 para D_2. Estos coeficientes se traducen en tres lineas paralelas. El grupo de referencia (en este caso el grupo 3) se define por la ecuación de regresión con las dos variables ficticias igual a cero. La línea del grupo 1 está dos unidades por encima de la línea para el grupo de referencia. La línea del grupo 2 está tres unidades por debajo de la línea del grupo de referencia 3. Las líneas paralelas indican que las variables ficticias



Ecuaciones de regresión con variables ficticias (D_1 y D_2)	
Especificadas	$Y = a + b_1X + b_2D_1 + b_3D_2$
Estimadas	
Globales	$Y = 2 + 1{,}2X + 2D_1 - 3D_2$
Específicas de grupo	
Grupo 1 ($D_1 = 1, D_2 = 0$)	$Y = 2 + 1{,}2X + 2(1)$
Grupo 2 ($D_1 = 0, D_2 = 1$)	$Y = 2 + 1{,}2X - 3(1)$
Grupo 3 ($D_1 = 0, D_2 = 0$)	$Y = 2 + 1{,}2X$

FIGURA 4.3. La incorporación de variables no métricas mediante variables ficticias.

No cambia la naturaleza de la relación, pero solamente estipula los atajos diferentes entre los grupos. Esta forma es más apropiada cuando existe un grupo de comparación lógica, tal como ocurre en un experimento. Siempre que se utilice la codificación mediante variable ficticia, debemos ser conscientes del grupo de comparación y recordar que los coeficientes presentan las diferencias entre la media del grupo y la del grupo de referencia.

Un método alternativo de codificación de variable ficticia es la **codificación de efectos**. Es exactamente igual que la codificación de indicador excepto que el grupo omitido o de comparación (el grupo donde todos son ceros) se le da ahora el valor de –1 en lugar de 0 para las variables ficticias. Así, los coeficientes representan diferencias para cualquier grupo respecto de la media de todos los grupos (en vez de la media del grupo omitido). Ambas formas de variable ficticia darán exactamente los mismos resultados predictivos, coeficientes de determinación y coeficientes de regresión de las variables continuas. Las únicas diferencias estarán en la interpretación de los coeficientes de la variable ficticia.

Representación de efectos curvilíneos con polinomios

Existen varios tipos de transformaciones de datos apropiados para convertir en lineal una relación curvilínea. Los métodos directos, discutidos en el Capítulo 2, implican modificaciones de los va-

lores a través de ciertas transformaciones aritméticas (por ejemplo, calculando la raíz cuadrada o el logaritmo de las variables). Sin embargo, tales transformaciones tienen varias limitaciones. En primer lugar, son útiles sólo en relaciones curvilineales simples (una relación con sólo un punto de giro o inflexión). En segundo lugar, no ofrecen medios estadísticos para evaluar si el modelo lineal o curvilineal es el más apropiado. Finalmente, sólo se pueden utilizar para relaciones univariantes y no para la interacción entre variables cuando nos encontramos con más de una variable independiente. Discutiremos a continuación un medio de crear variables que modelizan explícitamente los componentes curvilineales de la relación y ponen de manifiesto las limitaciones inherentes a las transformaciones de los datos.

Los **polinomios** son transformaciones potenciales de una variable independiente que añaden una componente no lineal para cada potencia adicional de la variable independiente. La potencia de 1 (X^1) representa el componente lineal y es la forma que vamos a discutir a lo largo de este capítulo. La potencia segunda, la variable al cuadrado (X^2), representa el componente cuadrático. En términos gráficos, X^2 representa el primer punto de inflexión. Un componente cúbico, representado por la variable elevada al cuadrado (X^3), añade un segundo punto de inflexión. Con estas variables e incluso con potencias superiores, pueden incluirse relaciones más complejas de las que son posibles explicar sólo con transformaciones. Por ejemplo, en un modelo de regresión simple, un modelo curvilineal con un punto de giro puede modelizarse con la ecuación:

$$Y = b_0 + b_1X_1 + b_2X_1^2$$

donde

b_0 = constante
b_1X_1 = efecto lineal de X_1
$b_2X_1^2$ = efecto curvilineal de X_1

Aunque puede añadirse cualquier número de componentes no lineales, el término cúbico es la mayor potencia utilizada habitualmente. A medida que cada nueva variable entra en la regresión, podemos realizar un test estadístico directo de los componentes no lineales que no podemos hacer con las transformaciones de los datos. En la Figura 4.4 se muestran tres relaciones (dos no lineales y una lineal). A efectos interpretativos, el término cuadrático positivo indica una curva en forma de $\cup$ hacia arriba, mientras que un coeficiente negativo indica una relación con la $\cap$ hacia abajo.

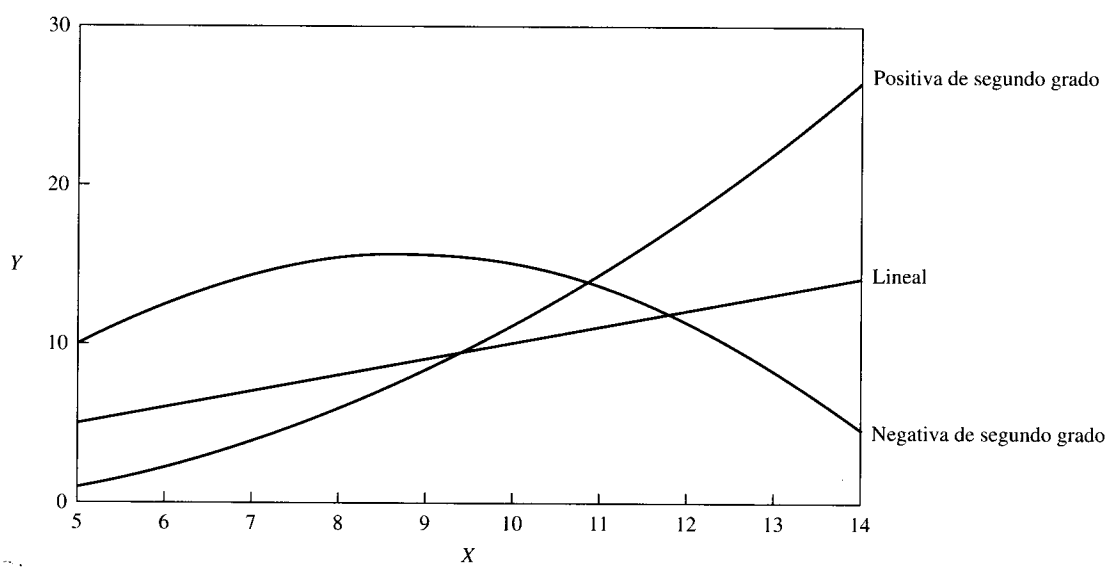


FIGURA 4.4. Representación de las relaciones no lineales con polinomios.

Los polinomios multivariantes se crean cuando la ecuación de la regresión contiene dos o más variables independientes. Seguimos el mismo procedimiento para crear los términos del polinomio al igual que antes, pero ahora debemos crear además un término adicional, el término de interacción (X_1X_2), que se necesita para cada combinación de variables para representar completamente los efectos multivariantes. En términos gráficos, un polinomio multivariante de dos variables se representa como una superficie con un pico o un valle. Para polinomios de orden superior, es mejor hacer la interpretación dibujando la superficie desde los valores previstos.

¿Cuántos términos deberían añadirse? Una práctica común es empezar con el componente lineal y entonces añadir secuencialmente polinomios de orden superior hasta que se llegue a la no significación. El uso de los polinomios no está exento de problemas potenciales. En primer lugar, cada término adicional requiere un mayor grado de libertad, que puede ser particularmente restrictivo con tamaños muestrales pequeños. Esta limitación no ocurre con las transformaciones de los datos. Asimismo, se introduce la multicolinealidad en los términos adicionales y hace que la comprobación de la significación estadística de los términos de los polinomios sea inapropiada. En su lugar, el investigador tiene que comparar los valores R^2 del modelo de ecuación con los términos lineales con la R^2 para la ecuación con los términos de los polinomios. Las pruebas para la significación estadística del R^2 en aumento representan la manera apropiada de valorar el impacto de los polinomios.

Representación de la interacción o efectos moderadores

Las relaciones no lineales arriba mencionadas requieren la creación de una variable adicional (por ejemplo, el término al cuadrado) para representar un cambio de pendiente en la relación sobre el rango de la variable independiente. Ésta se centra en la relación entre una única variable independiente y la variable dependiente. Pero ¿qué ocurre si una relación de variable dependiente/independiente se ve afectada por otra variable independiente? A esto se le llama el **efecto moderador**, que ocurre cuando la variable moderador, una segunda variable independiente, cambia la *forma* de la relación entre otra variable independiente y la variable criterio. También se conoce como un *efecto interacción* y es similar al término interacción que se encuentra en el análisis de la varianza y el análisis multivariante de la varianza (véase Capítulo 6 para más detalles acerca de los términos de interacción).

El efecto moderador más comúnmente empleado en la regresión múltiple es el *moderador cuasi* o *bilineal*, donde la pendiente de la relación de una variable independiente (X_1) cambia junto con los valores de la variable moderador (X_2) [7, 14]. En nuestro ejemplo anterior del uso de tarjetas de crédito, supongamos que la renta familiar (X_2) se encontró que era un moderador positivo de la relación entre el tamaño de la familia (X_1) y el uso de tarjetas de crédito (Y). Esto significaría que el cambio esperado en el uso de tarjetas de crédito basado en el tamaño de la familia (b_1, el coeficiente de regresión de X_1) podría ser menor para familias con rentas bajas y mayor para familias con rentas altas. Sin el efecto moderador, suponemos que el tamaño de la familia tiene un efecto «constante» sobre el número de tarjetas de crédito utilizadas. Pero los términos de interacción nos dicen que esta relación cambia, según el nivel de renta de la familia. Nótese que esto no significa necesariamente que los efectos del tamaño o la renta familiar por sí mismos no sean importantes, sino que el término de interacción complementa su explicación del uso de tarjetas de crédito.

El efecto moderador se representa en la regresión múltiple por un término bastante similar a los polinomios descritos previamente para representar efectos no lineales. El término moderador es una variable compuesta formada por la multiplicación de X_1 por el moderador X_2, que entra en la ecuación de regresión. De hecho, el término no lineal puede ser considerado como una forma de interacción, donde la variable independiente se «modera» a sí misma, elevándose al cuadrado (X_iX_i). La relación moderadora se representa como:

$$Y_0 = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2$$

donde

b_0 = constante
b_1X_1 = efecto lineal de X_1
b_2X_2 = efecto lineal de X_2
$b_3X_1X_2$ = efecto moderador de X_2 sobre X_1

Dada la multicolinealidad entre las variables antiguas y nuevas, se emplea un enfoque parecido a la comprobación para la significación de efectos polinomiales (no lineales). Para determinar si el efecto moderador es significativo, el investigador debe estimar en primer lugar la ecuación original (sin moderar) y a continuación estima la relación moderada. Si el cambio en el R^2 es estadísticamente significativo, entonces nos hallamos en presencia de un efecto moderador significativo. Por lo que solamente se valora el efecto de incremento, no las variables individuales.

La interpretación de los coeficientes de regresión cambia ligeramente en las relaciones moderadas. El coeficiente b_3, el efecto moderador, indica que el cambio unitario en el efecto de X_1 cuando X_2 cambia. Los coeficientes b_1 y b_2 representan los efectos de X_1 y X_2, respectivamente, cuando el resto de las variables independientes es cero. En la relación sin moderar, el coeficiente b_1 representa el efecto de X_1 para todos los niveles de X_2 y viceversa para b_2. Por tanto, en la regresión sin moderar, los coeficientes de regresión b_1 y b_2 se «promedian» respecto de los niveles del resto de las otras variables independientes, mientras que en una relación moderada se separan del resto de las variables independientes. Para determinar el efecto total de una variable independiente, se deben combinar los efectos separados y moderados. El efecto total conjunto de X_1 para cualquier valor de X_2 se puede calcular sustituyendo el valor de X_2 en lo siguiente:

$$b_{(\text{total})} = b_1 + b_3X_2$$

Por ejemplo, supongamos una regresión moderada que se resuelve en los siguientes coeficientes: $b_1 = 2{,}0$ y $b_3 = 0{,}5$. Si el valor de X_2 puede tomar valores entre uno y siete, el investigador puede calcular el efecto total de X_1 para cualquier valor de X_2. Cuando X_2 es igual a 3, el efecto total de X_1 es 3,5 [2,0 + 0,5(3)]. Cuando X_2 llega a 7, el efecto total de X_1 es ahora 5,5 [2,0 + 0,5(7)]. Podemos ver en acción al efecto moderador, haciendo que la relación de X_1 y la variable criterio cambie, dado el nivel de X_2. Podemos encontrar una excelente discusión de las relaciones moderadas en varias fuentes [5, 7, 14].

Resumen

La creación de nuevas variables proporciona al investigador una flexibilidad enorme en la representación de una amplia gama de relaciones dentro de los modelos de regresión. No obstante, demasiadas veces el deseo de obtener un ajuste mejor de un modelo lleva a la inclusión de estas relaciones especiales sin apoyo teórico. En estos casos, el investigador corre un riesgo mayor de encontrar resultados con poca o ninguna generalización. En su lugar, con el empleo de estas variables adicionales, el investigador tiene que estar guiado por una teoría respaldada por el análisis empírico. De esta manera, se puede alcanzar tanto la significación práctica como estadística.

Tercer paso: Supuestos en el análisis de regresión múltiple

Hemos mostrado cómo son posibles mejoras en la predicción de la variable criterio añadiendo variables independientes e incluso transformándolas para representar aspectos de las relación que no son lineales. Pero para hacerlo debemos hacer varios supuestos sobre las relaciones entre las variables dependientes e independientes que afectan al procedimiento estadístico (mínimos cuadrados) utilizado para la regresión múltiple. En las siguientes secciones discutiremos la contrastación de los supuestos y las acciones correctivas que se deben tomar si se incumplen los resultados.

Valoración de las variables individuales frente al valor teórico

Los supuestos subyacentes de la regresión múltiple se aplican tanto a las variables individuales (dependientes e independientes) como a la relación global. En el Capítulo 2 se examinaron los métodos disponibles para evaluar los supuestos de las variables individuales. Pero en la regresión múltiple, una vez que se ha calculado el valor teórico, actúan colectivamente en la predicción de la variable dependiente. Este hecho implica que se deben evaluar los supuestos no sólo de las variables individuales sino del valor teórico en sí mismo. Esta sección se centra en el examen del valor teórico y de su relación con la variable dependiente en el cumplimiento de los supuestos de la regresión múltiple. Estos análisis deben realizarse *después* de que se haya estimado el modelo de regresión en el paso cuarto. Por tanto, la contrastación de los supuestos debe tener lugar no sólo en las fases iniciales de la regresión sino también después de que el modelo se ha estimado.

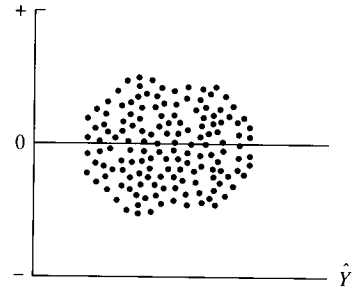
La cuestión básica es si, en el proceso de cálculo de los coeficientes de regresión y predicción de la variable criterio, se cumplen los supuestos del análisis de la regresión. ¿Son los errores en la predicción un resultado de una falta efectiva de relación entre variables o son provocados por ciertas características de los datos no contemplados por el modelo de regresión? Los supuestos que se van a examinar son los siguientes:

- La linealidad del fenómeno medido
- La varianza constante del término de error
- La independencia de los términos de error
- La normalidad de la distribución del término de error

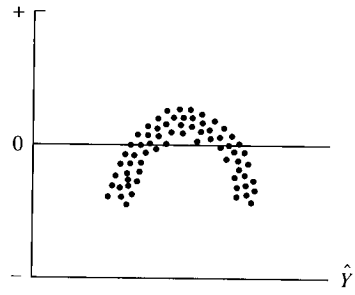
La medida principal del error de predicción del valor teórico es el **residuo** —la diferencia entre los valores observados y las predicciones de la variable criterio. Los gráficos de residuos y de las variables independientes o de las predicciones constituyen el método básico de identificación de los incumplimientos de los supuestos para el conjunto de la relación. Cuando se examinan los residuos, se recomienda cierta forma de estandarización, con el fin de hacer los residuos directamente comparables. (En su forma original, los valores con sobrepredicción tienen mayores residuos.) El método más ampliamente utilizado es el **residuo basado en la *t* de Student**.

Su valor corresponde a valores *t*. Esta correspondencia facilita la evaluación de la significación estadística de residuos particularmente elevados. En las siguientes secciones, examinaremos también unas series de test estadísticos que pueden complementar el examen de los gráficos de residuos.

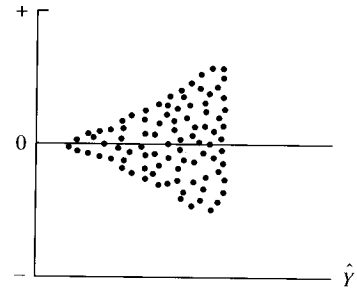
El gráfico de residuos más habitual se forma con los residuos (r_i) frente a los valores de la predicción de la dependiente (Y_i). Para un modelo de regresión simple, se pueden trazar los residuos respecto de las variables dependientes o independientes, dado que están directamente relacionados. En la regresión múltiple, sin embargo, sólo los valores dependientes pronosticados representan el efecto total del valor teórico de regresión. Por tanto, a no ser que el análisis residual pretenda concentrarse en una sola variable, se usan las variables dependientes pronosticadas. Los incumplimientos de cada supuesto pueden identificarse por específicas regularidades de los residuos. La Figura 4.5 contiene unos cuantos gráficos de residuos que muestran los supuestos básicos discutidos en la secciones siguientes. Un gráfico de especial interés es el **gráfico de no-correlación de residuos** (Figura 4.5a), el gráfico de los residuos cuando se cumplen todos los supuestos. Los gráficos de no-correlación de los residuos se distribuyen aleatoriamente, con una dispersión relativamente igual a cero y una tendencia no muy fuerte a que sea mayor o menor que cero. Asimismo, no se encuentra ninguna pauta o regularidad para valores elevados o reducidos de la variable independiente. Los gráficos residuales restantes serán utilizados para ilustrar los métodos de examen de incumplimientos de los supuestos que subyacen el análisis de regresión.



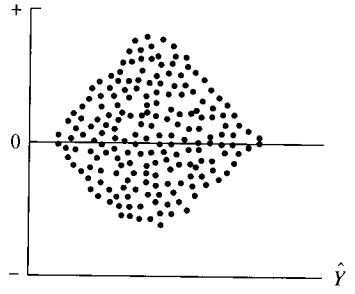
(a) Gráfico de no correlación de residuos



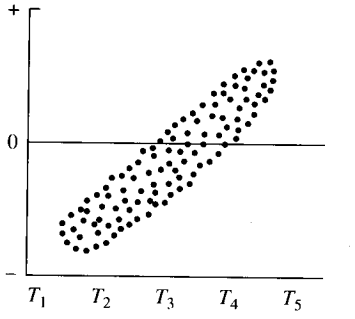
(b) No linealidad



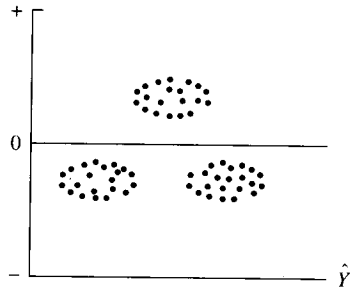
(c) Heterocedasticidad



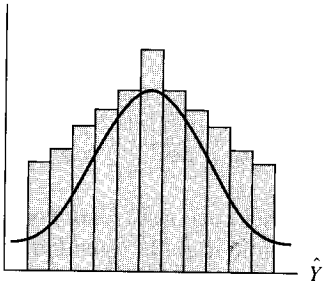
(d) Heterocedasticidad



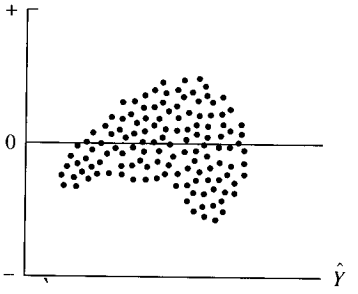
(e) Dependencia temporal



(f) Dependencia de evento



(g) Histograma normal



(h) No linealidad y heterocedasticidad

FIGURA 4.5. Análisis gráfico de los residuos.

Linealidad del fenómeno

La **linealidad** de la relación entre variables dependientes o independientes representa el grado de cambio en la variable dependiente asociado con la variable independiente. El coeficiente de regresión es constante a lo largo del rango de valores de la variable independiente. El concepto de correlación está basado en una relación lineal, siendo por tanto un supuesto crítico del análisis de regresión. La linealidad se puede examinar fácilmente en los gráficos de residuos. La Figura 4.5b muestra una forma típica de residuos que indican la existencia de una relación no lineal no representada en el modelo habitual. Cualquier modelo curvilineal de los residuos indica que la acción correctiva aumentará tanto la precisión predictiva del modelo como la validez de los coeficientes estimados. Las soluciones mediante transformaciones de los datos se discutieron en el Capítulo 2 [10]. El investigador puede también querer incluir relaciones no lineales en nuestros modelos de regresión. En el segundo paso se presentan las transformaciones de datos, tales como la creación

de términos polinomiales, o bien métodos específicos como la regresión no lineal que pueden tratar los efectos curvilineales de las variables independientes o relaciones no lineales más complejas.

En la regresión múltiple con más de una variable independiente, el examen de los residuos mostraría los efectos combinados de todas las variables independientes, pero no podemos examinar el efecto de cualquier variable independiente separadamente en un gráfico de residuos. Para hacerlo utilizamos lo que se denomina **gráfico de regresión parcial**, que muestra la relación de una única variable independiente con relación a la variable dependiente. Difiere del gráfico de residuos que acabamos de discutir en que la línea que atraviesa la nube de puntos, que era horizontal en los gráficos anteriores (véase Figura 4.5), tendrá ahora pendiente positiva o negativa dependiendo de si el coeficiente de regresión para esa variable independiente es positivo o negativo. El examen de los residuos alrededor de esta línea se hace exactamente igual que antes.

En los gráficos de regresión parciales, las pautas curvilineales de los residuos indican una relación no lineal entre una variable independiente y la variable dependiente. Este es el método más útil cuando tenemos varias variables independientes, en la medida en que podemos decir qué variables específicas incumplen el supuesto de linealidad y aplicar los remedios necesarios. También se facilita la identificación de atípicos o de observaciones influyentes sobre la base del uso de una sola variable independiente a la vez.

Varianza constante del término de error

La presencia de varianzas desiguales (**heterocedasticidad**) es uno de los supuestos que se incumple más habitualmente. El diagnóstico se realiza mediante gráficos de residuos o test estadísticos simples. El gráfico de los residuos (basados en la *t* de Student) frente a los valores de la variable dependiente se compone con el gráfico de no-correlación de residuos (véase Figura 4.5a) y muestra una forma consistente si la varianza no es constante. Quizá la forma más común es la triangular, en cualquier dirección (Figura 4.5c). Puede esperarse una forma de diamante (Figura 4.5d) en el caso donde se espera más variación en los valores intermedios que en los extremos. Muchas veces, muchos incumplimientos ocurren simultáneamente, tal como la no-linealidad y la heterocedasticidad mostrada en la Figura 4.5h. Las soluciones para cada uno de los problemas de incumplimiento a menudo corrigen otros problemas.

Cada programa estadístico de computador tiene test estadísticos de heterocedasticidad. Por ejemplo, el SPSS ofrece el test de Levane de homogeneidad de varianza, que mide la igualdad de varianzas para un único par de varianzàs. Su uso es particularmente recomendable porque es el que menos queda afectado por desviaciones de la normalidad, otro de los problemas que ocurren con frecuencia en la regresión.

Si existe heterocedasticidad, podemos aplicar dos soluciones. Si el incumplimiento puede atribuirse a una única variable criterio, puede utilizarse el procedimiento de mínimos cuadrados ponderados. Más directas y fáciles son, sin embargo, diversas transformaciones de estabilización de varianza discutidas en el Capítulo 2, que permiten la utilización de las variables transformadas en nuestro modelo de regresión.

Independencia de los términos de error

En la regresión suponemos que cada variable predictor es independiente. Con esto queremos decir que el valor de la predicción no está relacionado con cualquier otra predicción; esto es, no están ordenadas por otra variable. Para identificar este hecho, se utiliza el gráfico de residuos respecto a cualquier posible varianza secuencial. Si los residuos son independientes, la forma puede parecer aleatoria y similar al gráfico de no-correlación de los residuos. Los incumplimientos quedarán identificados por una forma consistente de los residuos. La Figura 4.5e representa un gráfico de residuos que muestra una asociación entre los residuos y tiempo, una variable de secuencia habitual. Otra forma habitual se muestra en la Figura 4.5f. Esta forma ocurre cuando las condiciones

básicas del modelo cambian pero no se incluyen en el modelo. Por ejemplo, las ventas de trajes de baño se miden mensualmente en 12 meses, con dos estaciones invernales frente a una única estación de verano, sin estimarse un indicador estacional. La forma de los residuos mostrará residuos negativos para los meses de invierno frente a residuos positivos para los meses de verano. Las transformaciones de los datos, tales como primeras diferencias en un modelo de series temporales, inclusión de variables indicador o modelos de regresión especialmente formulados pueden solucionar este problema.

Normalidad de la distribución del término de error

Quizá el incumplimiento de supuestos más frecuente es la no-normalidad de las variables independientes y dependientes, o ambos [13]. El diagnóstico más simple para el conjunto de variables predictor en la ecuación es un histograma de residuos, donde se puede comprobar visualmente si la distribución se aproxima a la normal (véase Figura 4.5g). Aunque atractivo por su simplicidad, este método es particularmente difícil en muestras pequeñas, donde la distribución puede estar deformada. Un método mejor es utilizar los **gráficos de probabilidad normal**. Difieren de los gráficos de residuos en que los residuos estandarizados se comparan con la distribución normal. La distribución normal traza una línea recta diagonal y los gráficos de residuos se comparan con la diagonal. Si una distribución es normal, la línea de residuos seguirá de cerca la diagonal. Con el mismo procedimiento se pueden comparar las variables dependientes e independientes separadamente respecto de la distribución normal [6]. El Capítulo 2 da una mayor discusión de la interpretación de gráficos de probabilidad normal.

Resumen

El análisis de residuos, bien con los gráficos de residuos o bien con test estadísticos, proporciona un conjunto simple pero potente de instrumentos analíticos para examinar la conveniencia de nuestro modelo de regresión. Demasiadas veces, sin embargo, estos análisis no se realizan y se dejan intactos los incumplimientos de los supuestos. Así, los usuarios no son conscientes de las potenciales imprecisiones que puedan presentar. Éstas van desde test de significación de coeficientes inapropiados (por mostrar significación cuando no está presente o viceversa) a las predicciones sesgadas e imprecisas de la variable dependiente. Recomendamos fehacientemente que estos métodos se apliquen para conjunto de datos del modelo de regresión. La aplicación de los remedios, particularmente la transformación de los datos, aumentará la confianza en las interpretaciones y predicciones de la regresión múltiple.

Cuarto paso: Estimación del modelo de regresión y valoración

Habiendo especificado los objetivos del análisis de regresión, seleccionadas las variables dependientes e independientes, enfrentados a los resultados del diseño de la investigación y evaluadas las variables a la hora de cumplir los supuestos de la regresión, estamos preparados para estimar el modelo de regresión y evaluar la precisión predictiva conjunta de las variables independientes (véase Figura 4.6). A este nivel, se deben lograr tres tareas básicas: (1) seleccionar un método para especificar el modelo de regresión a estimar, (2) evaluar la significación estadística del modelo conjunto en la predicción de la variable criterio y (3) determinar si cualquiera de las observaciones ejerce una indebida influencia sobre los resultados.

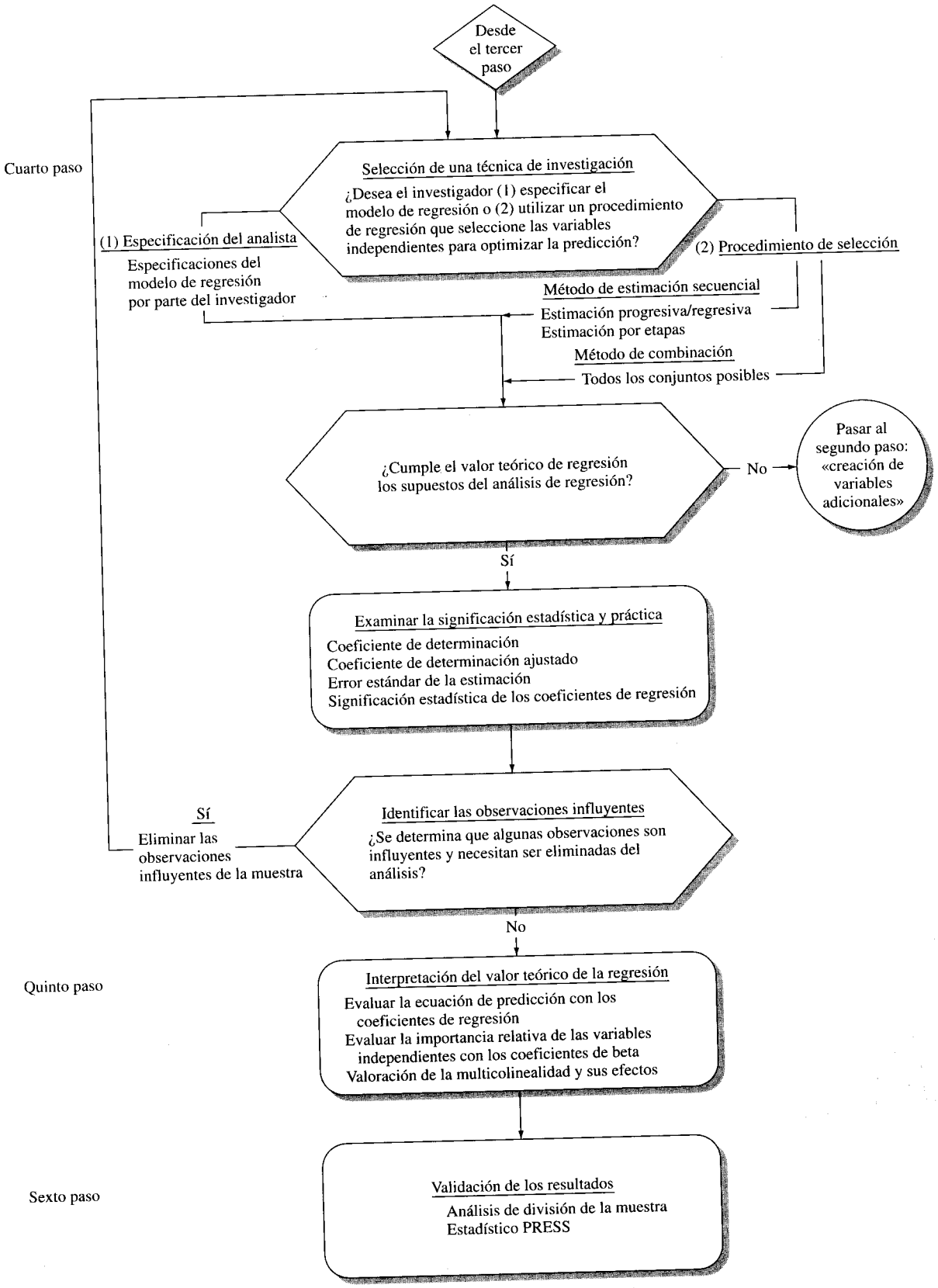


FIGURA 4.6. Pasos 4-6 en diagrama de decisión de la regresión múltiple.

Aproximaciones generales a la selección de variables

En la mayoría de los casos de regresión múltiple, se tiene un número posible de variables independientes entre las cuales elegir para incluirlas en la ecuación de regresión. A veces el conjunto de variables independientes puede estar muy definido y el modelo de regresión se usa esencialmente en una aproximación confirmatoria. En otros casos, el investigador puede desear elegir entre el conjunto de variables independientes. Existen varias aproximaciones (métodos de búsqueda secuencial y procesos combinatorios) para ayudar al investigador en la búsqueda del «mejor» modelo de regresión. A continuación se discutirá cada uno de estos modelos de especificación del modelo de regresión.

Especificación confirmatoria

La más simple aunque quizá más exigente aproximación de especificación del modelo de regresión es emplear una perspectiva confirmatoria en la cual el investigador especificaría por completo el conjunto de variables independientes a incluir. En comparación con las aproximaciones que se discutirán a continuación, se tiene control total sobre la selección de la variable. Aunque la especificación confirmatoria es simple en su concepción, debemos estar seguros de que el conjunto de variables consigue la máxima predicción mientras mantiene un modelo de parsimonia. Las pautas generales para el desarrollo del modelo se discuten en los Capítulos 1 al 11.

Métodos de búsqueda secuencial

Los métodos de búsqueda secuencial tienen en común la aproximación general de estimación de las ecuaciones de regresión con un conjunto de variables y a continuación añadir o eliminar selectivamente variables hasta que se consiga alguna medida criterio conjunta. Esta aproximación proporciona un método objetivo de selección de variables que maximizan la predicción con el número más pequeño de variables empleadas. Existen dos tipos de aproximaciones de búsqueda secuencial: (1) la estimación por etapas, y (2) la eliminación progresiva y regresiva. En cada aproximación, se valoran las variables individualmente en función de su contribución a la predicción de la variable dependiente y se añaden o eliminan según su contribución relativa.

Estimación por etapas (paso a paso o stepwise) La **estimación por etapas** es quizá la aproximación más popular para seleccionar variables. Esta aproximación permite examinar la contribución de cada variable predictor al modelo de regresión. Se considera la inclusión de cada variable antes de desarrollar la ecuación. Se añade primero la variable independiente con la contribución más grande. Las variables independientes se seleccionan entonces para la inclusión basada en su contribución incremental sobre la(s) variable(s) ya existente(s) en la ecuación. Se ilustra el procedimiento por etapas en la Figura 4.7. Las cuestiones específicas en cada etapa son:

1. Empezar con el modelo de regresión simple en el cual sólo se utiliza la única variable predictor que es la que se muestra más altamente correlacionada con la variable criterio. La ecuación sería $Y = b_0 + b_1X_1$.
2. Examinar los **coeficientes de correlación parcial** para encontrar una variable predictor adicional que explique además de una parte significativa, la mayor parte del error que queda de la primera ecuación de regresión.
3. Recalcular la ecuación de regresión utilizando las dos variables predictor, y examinar el **valor parcial *F*** de la variable original del modelo para ver si todavía realiza una contribución significativa, dada la presencia de la nueva variable predictor. Si no lo hace, eliminamos la variable. Esta capacidad de eliminar variables presentes en el modelo distingue el modelo por etapas de los modelos de adición progresiva/eliminación regresiva. Si las variables originales todavía representan una contribución significativa, la ecuación sería $Y = b_0 + b_1X_1 + b_2X_2$.

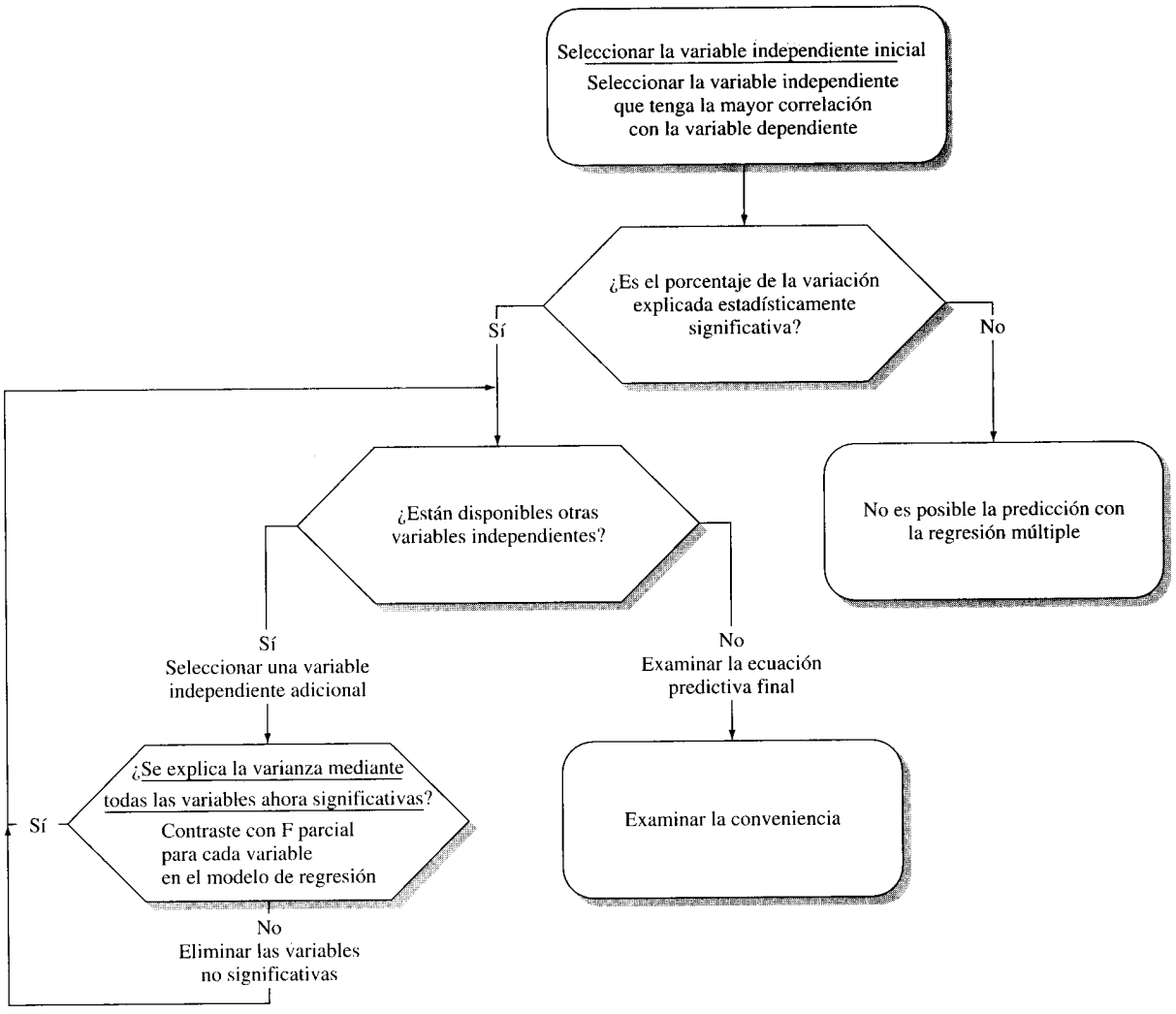


FIGURA 4.7. Diagrama de flujos del procedimiento de estimación por etapas.

4. Continúa este procedimiento examinando todos las variables independientes no presentes en el modelo para determinar si deberían incluirse en la ecuación. Si se incluye una nueva variable independiente, hay que examinar todos los predictores previamente incluidos en el modelo para juzgar si se deben mantener. Existe un sesgo potencial en el procedimiento por etapas que resulta de considerar sólo una variable a seleccionar cada vez. Supongamos que las variables X_3 y X_4 explicaran conjuntamente una parte significativa de la varianza (cada una considerando la presencia de la otra), pero no son significativas por sí solas. En esta situación, ninguna debería ser considerada para el modelo final.

La adición progresiva (forward) y la eliminación regresiva (backward) **La adición progresiva** y **la eliminación regresiva** son fundamentalmente procesos de ensayo y error para buscar los mejores estimadores de la regresión. El modelo de adición progresiva es similar al procedimiento por etapas arriba explicada, mientras que el procedimiento de eliminación regresiva implica calcular una ecuación de regresión con todas las variables independientes, para a continuación ir eliminando las variables independientes que no contribuyan significativamente. La distinción principal de la aproximación por etapas respecto de los procedimientos de adición progresiva y eliminación regresiva es su capacidad de añadir o eliminar las variables en cada etapa. Una vez que se añade o elimina una variable en las esquemas de adición progresiva o eliminación regresiva, no existe posibilidad de revertir la acción posteriormente.

Advertencias sobre los métodos de búsqueda secuenciales El investigador debe ser consciente de dos advertencias cuando se usa cualquier procedimiento de búsqueda secuencial. En primer lugar, la multicolinealidad entre variables independientes puede tener un impacto sustancial sobre la especificación final del modelo. Examinemos esta situación con dos variables altamente correlacionadas que tienen similares correlaciones con la variable independiente. El criterio de inclusión o eliminación en estas aproximaciones es maximizar el incremento de potencia predictiva de la variable adicional. Si una de estas variables entra en el modelo de regresión, es muy probable que la otra variable también entre, dado que estas variables están altamente correlacionadas y existe poca varianza singular para cada variable por separado (véase el debate más adelante sobre la multicolinealidad).

Por esta razón, se deben evaluar los efectos de la multicolinealidad en la interpretación del modelo y examinar las correlaciones directas de todas las variables independientes potenciales. Esto ayudará a evitar concluir que las variables independientes que no entren en el modelo no sean trascendentes cuando en realidad están altamente relacionadas con la variable dependiente, pero también correlacionadas con las variables ya existentes en el modelo. Aunque las aproximaciones de búsqueda secuencial maximizarán la capacidad predictiva del modelo de regresión, el investigador debe ser cuidadoso en la interpretación del modelo. Una segunda advertencia pertenece principalmente al procedimiento por etapas. En esta aproximación, los test de significación múltiple se realizan en el proceso de estimación del modelo. Para asegurar que la tasa de error conjunto a lo largo de todos los test de significación es razonable, el investigador debería emplear umbrales muy conservadores (por ejemplo, 0,01) al añadir o destruir las variables.

Métodos combinatorios

Los métodos combinatorios son fundamentalmente un proceso de búsqueda generalizada a lo largo de todas las combinaciones posibles de variables independientes. El procedimiento más conocido es la **regresión parcial combinando variables**, que es exactamente lo que su nombre indica. Se examinan todas las combinaciones posibles de las variables independientes para identificar el conjunto de variables que mejor se ajusta. Por ejemplo, en un modelo con diez variables independientes, existen 1.204 regresiones posibles (una ecuación con una única constante, 10 ecuaciones con una única variable independiente, 45 ecuaciones con todas las combinaciones posibles de dos variables, etc.). Con procedimientos de estimación informáticos, este proceso se puede gestionar incluso para problemas muy grandes, identificando la mejor ecuación de regresión conjunta para cualquier número de medidas de ajuste predictivo. El investigador debe recordar que supuestos tales como la multicolinealidad, la identificación de atípicos y observaciones influyentes y la interpretación de los resultados no están orientadas a la selección del modelo final. Cuando se han considerado estos supuestos, la «mejor» ecuación puede tener problemas serios que afecten a su conveniencia, pudiendo ser elegido en última instancia otro modelo.

Perspectiva de las aproximaciones de la selección de modelos

Independientemente de que se elija un método combinacional, de búsqueda secuencial o confirmatorio, el criterio más importante es el conocimiento sustantivo del investigador de la situación, que es lo que determina las variables que se van incluir así como los signos esperados y la magnitud de sus coeficientes. Sin este conocimiento, la regresión resultante puede tener una elevada precisión predictiva sin relevancia teórica o gerencial. El investigador no debería guiarse completamente por estos métodos sino que en su lugar debería utilizarlos después de una cuidadosa consideración de las aproximaciones alternativas para a continuación aceptar los resultados sólo después de un cuidadoso escrutinio.

Contrastación del cumplimiento de los supuestos de regresión

Con las variables independientes seleccionadas y los coeficientes de regresión estimados, el investigador debe ahora evaluar el modelo estimado a la hora de cumplir los supuestos subyacentes en la regresión múltiple. Como se discutió en el paso tercero, las variables individuales deben cumplir los supuestos de linealidad, varianza constante, independencia y normalidad. Además de las variables individuales, el valor teórico de la regresión debe también cumplir estos supuestos. Los test de diagnóstico expuestos en el paso tercero pueden aplicarse a la evaluación del efecto colectivo del valor teórico a través del examen de los residuos. Si se encuentran incumplimientos sustanciales, el investigador debe tomar medidas correctivas para posteriormente volver a estimar el modelo de regresión.

Examen de la significación estadística de nuestro modelo

Si fuéramos a tomar muestras repetidas de ocho familias y preguntáramos cuántos miembros de la familia y tarjetas de crédito tienen, rara vez obtendríamos exactamente los mismos valores para $Y = b_0 + b_1X_1$ para todas las muestras. Esperaríamos variaciones al azar causadas por las diferencias entre las muestras. Generalmente tomaríamos sólo una muestra y basaríamos sobre ella nuestro modelo predictivo. Con una sola muestra y modelo de regresión, necesitamos comprobar la hipótesis con relación a nuestro modelo predictivo para asegurar que representa la población de todas las familias que tienen tarjetas de crédito en lugar de representar sólo a nuestra muestra de ocho familias. Estos test pueden tomar una o dos formas básicas: un test de varianza explicada (coeficiente de determinación) y los test de coeficientes.

Significación del modelo en su conjunto: El coeficiente de determinación

Para contrastar la hipótesis de que la cantidad de variación explicada por el modelo de regresión es más que la variación explicada por la media (es decir, que R^2 es mayor que cero), se utiliza el ratio F. La prueba del estadístico de la F se define como:

$$\text{Ratio } F = \frac{\dfrac{\text{Suma de los errores al cuadrado}_{\text{regresión}}}{\text{Grados de libertad}_{\text{regresión}}}}{\dfrac{\text{Suma de los errores al cuadrado}_{\text{total}}}{\text{Grados de libertad}_{\text{residuos}}}} = \frac{SSE_{\text{regresión}}/df_{\text{regresión}}}{SSE_{\text{total}}/df_{\text{residual}}}$$

donde

$\text{Grados de libertad}_{\text{regresión}}$ = Número de los coeficientes estimados (incluida la constante) – 1.
$\text{Grados de libertad}_{\text{residuos}}$ = Tamaño muestral – el número de los coeficientes estimados (incluida la constante).

Deben destacarse dos importantes características de este ratio:

1. Cada suma de los cuadrados se divide por sus **grados de libertad**. Es simplemente la varianza de los errores de predicción.
2. Intuitivamente, si el ratio de la varianza explicada respecto a la varianza sobre la línea básica (alrededor de la media) es alto, el valor teórico de regresión tiene que tener un valor significativo en la explicación de la variable dependiente.

En nuestro ejemplo, el ratio F del modelo de regresión simple discutido previamente en el capítulo es $(16{,}5 \div 1)/(5{,}50 \div 6) = 18{,}0$. La tabla del estadístico de la F de 1 con seis grados de libertad para un nivel de significación de 0,05 proporciona 5,99. Dado que el ratio F es mayor que el valor de tabla, rechazamos la hipótesis de que la reducción en el error que obtenemos utilizando el tamaño de la familia para predecir la posesión de tarjetas de crédito era un producto del azar. Este resultado significa que, considerando la muestra utilizada para la estimación, podemos explicar 18 veces más variación que cuando utilizamos la media, y que esto no es muy probable que ocurra por azar (menos del 5 por ciento de las veces). De la misma forma, el ratio F del modelo de regresión múltiple con dos variables independientes es $(18{,}96 \div 2)/(3{,}04 \div 5) = 15{,}59$. El modelo de regresión múltiple es también estadísticamente significativo, indicando que la variable independiente adicional era sustancial al añadirse a la capacidad predictiva del modelo.

Sabemos que R^2 está influenciado por el número de variables predictor relativas al tamaño muestral. Se han propuesto varias reglas, que van desde 10 a 15 observaciones por predictor a un mínimo absoluto de 4 observaciones por variable independiente. A medida que se llega a estos límites, necesitamos ajustar la inflación del R^2 del «sobreajuste» de los datos. Como parte de todos los programas de regresión, se da un **coeficiente de regresión ajustado (R^2 ajustado)** junto con los coeficientes de determinación. Interpretado igualmente que el coeficiente de regresión sin ajustar, el R^2 ajustado se hace más pequeño a medida que tenemos menos observaciones por variable independiente. El R^2 ajustado es particularmente útil para comparar las diferentes ecuaciones de regresión estimadas con distintas variables independientes o diferentes tamaños muestrales, dado que marca límites para el número específico de variables independientes y para el tamaño muestral sobre el que se basa cada modelo. En nuestro ejemplo del uso de tarjetas de crédito, el R^2 para el modelo de regresión simple es de 0,751 y el R^2 ajustado es 0,709. Conforme añadimos la segunda variable independiente, el R^2 aumenta a 0,861, pero el R^2 ajustado sólo aumenta a 0,806. En los dos casos, el R^2 ajustado refleja el ratio descendiente de los coeficientes estimados al tamaño muestral y compensa un «sobre ajuste» de los datos.

Test de significación de los coeficientes de regresión

La prueba de significación estadística de los coeficientes estimados del análisis de regresión es apropiada y necesaria cuando el análisis se basa en una muestra de la población y no es un censo. Cuando utilizamos una muestra para estimar el modelo de regresión, el investigador no está interesado en la regresión estimada sólo para la muestra, sino en la generalización de los resultados para la población. Para cada muestra extraída de la población, se obtendrá un valor diferente. Para muestras pequeñas, los coeficientes estimados variarán ampliamente de muestra a muestra. Pero a medida que el tamaño muestral aumenta, las muestras se hacen más representativas de la población y la variación en los coeficientes estimados para estas muestras mayores se espera que sean más pequeñas. Esto es verdad hasta que se estima el análisis utilizando la población. En este caso, no hay necesidad para la significación estadística porque la «muestra» es igual a, y por tanto perfectamente representativa de la población. La variación esperada de los coeficientes estimados (tanto los coeficientes constantes como de regresión) se denomina el **error estándar** de los coeficientes.

La significación estadística de los coeficientes de regresión proporciona una estimación probabilística de fundamento estadístico sobre si los coeficientes estimados a lo largo de un gran número de muestras de un cierto tamaño serán diferentes de cero. Si el tamaño muestral es pequeño, la variación puede ser muy grande como para decir con el necesario grado de certeza (al que nos vamos a referir como el *nivel de significación*) que el coeficiente no es igual a cero. Sin embargo, si el tamaño muestral es grande, el test tiene una mayor precisión porque la variación en los coeficientes es menor. Muestras más grandes no garantizan que los coeficientes no sean iguales a cero, pero hacen los test más precisos.

Un ejemplo de la variación muestral en un coeficiente de regresión Para ilustrar este punto, se obtuvieron 20 muestras aleatorias de cada uno de los cuatro tamaños muestrales (10, 25, 50 y 100 encuestados) de una gran base de datos. Se realizó una regresión simple para cada mues-

tra y los coeficientes de regresión estimados se registran en la Tabla 4.8. Como se puede observar, la variación en los coeficientes estimados es mayor para muestras de 10 encuestados, que varían desde un coeficiente bajo 2,20 al más elevado de 6,06. A medida que el tamaño muestral aumenta de 25 a 50 encuestados, la variación de los coeficientes baja considerablemente. Finalmente, las muestras de 100 encuestados tienen un rango de casi la mitad que las muestras de 10 encuestados (2,10 frente a 3,86). Con esto podemos ver que la capacidad de los test de significación para determinar si el coeficiente es realmente mayor que cero es más precisa con tamaños muestrales superiores.

TABLA 4.8. Variación muestral para los coeficientes de regresión estimada

	Tamaño muestral			
Muestra	10	25	50	100
1	2,58	2,52	2,97	3,60
2	2,45	2,81	2,91	3,70
3	2,20	3,73	3,58	3,88
4	6,06	5,64	5,00	4,20
5	2,59	4,00	4,08	3,16
6	5,06	3,08	3,89	3,68
7	4,68	2,66	3,07	2,80
8	6,00	4,12	3,65	4,58
9	3,91	4,05	4,62	3,34
10	3,04	3,04	3,68	3,32
11	3,74	3,45	4,04	3,48
12	5,20	4,19	4,43	3,23
13	5,82	4,68	5,20	3,68
14	2,23	3,77	3,99	4,30
15	5,17	4,88	4,76	4,90
16	3,69	3,09	4,02	3,75
17	3,17	3,14	2,91	3,17
18	2,63	3,55	3,72	3,44
19	3,49	5,02	5,85	4,31
20	4,57	3,61	5,12	4,21
Mínimo	2,20	2,52	2,91	2,80
Máximo	6,06	5,64	5,85	4,90
Alcance	3,86	3,12	2,94	2,10
Desviación estándar	1,28	0,85	0,83	0,54

Contrastes de significación en el ejemplo de regresión simple Cuando abordamos el modelo de regresión simple para el ejemplo de uso de tarjetas de crédito, dijimos que la ecuación de regresión para el número de tarjetas de crédito es $Y = b_0 + b_i V_1 = 2{,}87 + 0{,}971$ (tamaño de la familia). Contrastaremos dos hipótesis para este modelo de regresión:

Hipótesis 1. El valor de la constante 2,87 se debe al error muestral, siendo cero el término constante real apropiado para la población.

Con esta hipótesis, estaríamos comprobando simplemente si el término constante debería considerarse apropiado para nuestro modelo predictivo. Si se encuentra que no es significativamente distinto de cero, supondríamos que el término constante no se utilizaría para propósitos predictivos. El contraste apropiado es el test de la *t*, que se encuentra habitualmente en la mayoría de los programas informáticos de regresión. El valor *t* de un coeficiente es el coeficiente dividido por el error estándar. Por ejemplo, un coeficiente de 2,5 con un error estándar de 0,5 tendría un valor *t* de 5,0. Para determinar si el coeficiente es diferente de cero de forma significativa, se compara el valor *t* calculado con el valor de la tabla para el tamaño muestral y el nivel de confianza seleccio-

nado. Si nuestro valor es mayor que el valor de la tabla, podemos estar seguros (en nuestro nivel de confianza) que el coeficiente tiene un efecto estadístico significativo en el valor teórico de regresión.

Desde un punto de vista práctico, este test rara vez es necesario. Si los datos utilizados para desarrollar el modelo no incluyeron ciertas observaciones con todos los medidos para cero, el término constante está «fuera» de los datos y actúa sólo para posicionar al modelo. No será necesario entonces contrastar el modelo.

Hipótesis 2. El coeficiente 0,971 indica que un aumento de una unidad en el tamaño familiar se asocia con un aumento en el número medio de tarjetas de crédito mantenidas por 0,971 y que este coeficiente también difiere significativamente de cero.

Si así ocurriera por un error muestral, concluiríamos que el tamaño de la familia no tiene impacto sobre el número de tarjetas de crédito mantenidas. Téngase en cuenta que este no es un test de un cierto valor exacto del coeficiente sino más bien de si debiera utilizarse. Otra vez, el test apropiado es el test de la *t*. El investigador debe recordar que el test estadístico de los coeficientes de regresión sirve para asegurar que para todas las posibles muestras que pudiesen extraerse, el coeficiente de regresión debería ser diferente de cero. En nuestro ejemplo, el error estándar del tamaño de familia en el modelo de regresión simple es 0,229. El valor calculado *t* es 4,25(0,971 ÷ 0,229), lo cual tiene una probabilidad de 0,005. Esto significa que podemos estar seguros con una alta nivel de exactitud (99,5 por ciento) que el coeficiente debería ser incluido en la ecuación de regresión.

Resumen Los contrastes de significación de los coeficientes de regresión proporcionan al investigador una valoración empírica de su «verdadero» impacto. Aunque no constituye una prueba de validez, determina si los impactos representados por los coeficientes son generalizables para otras muestras de esta población. Muchas veces los investigadores olvidan que los coeficientes estimados en el análisis de regresión son específicos para la muestra utilizada en la estimación. Representan las mejores estimaciones para aquella muestra de observaciones, pero como muestran los citados resultados, los coeficientes pueden variar notablemente de una muestra a otra. Esto indica la necesidad de una actuación coordinada para validar cualquier análisis de regresión sobre muestra(s) diferente(s). De esta manera, el investigador tiene que esperar que cambien los coeficientes, pero el intento es para demostrar que generalmente se mantiene la relación en otras muestras para que se pueda suponer que los resultados sean generalizables para cualquier muestra obtenida de la población.

Identificación de observaciones influyentes

Hasta ahora nos hemos centrado en la identificación de pautas generales en el conjunto de observaciones. Aquí desviaremos nuestra atención a las observaciones individuales, con el objetivo de encontrar las observaciones que caen fuera de las pautas generales del conjunto de datos o que ejercen una fuerte influencia en los resultados de la regresión. Recordemos que estas observaciones no son necesariamente malas en el sentido de que deban ser omitidas. En muchos casos representan los elementos diferenciadores del conjunto de datos. Sin embargo, debemos identificarlas y evaluar su impacto antes de empezar. La siguiente sección introduce el concepto de observaciones influyentes y su impacto potencial sobre los resultados de la regresión, mientras que el Apéndice 4A contiene una discusión más detallada de los procedimientos para identificar observaciones influyentes.

Las observaciones influyentes se clasifican en tres casos: atípicos, puntos de apalancamiento e influyentes. Los **atípicos** son observaciones que tienen grandes valores residuales y pueden identificarse sólo con respecto a un modelo de regresión específico. Los atípicos han sido tradicionalmente la única forma de observación influyente considerada en los modelos de regresión, y se han desarrollado métodos de regresión específicos (regresión robusta) para tratar específicamente con el impacto de atípicos sobre los resultados de la regresión [1, 12]. En el Capítulo 2 se ofrecieron procedimientos adicionales para identificar atípicos. Los **puntos de apalancamiento** son obser-

vaciones diferentes del resto de las observaciones de los valores de las variables independientes. Su impacto es particularmente destacado en los coeficientes estimados de una o más variables predictor. Finalmente, las **observaciones influyentes** en sentido amplio, incluyen todas las observaciones que tienen un efecto desproporcionado sobre los resultados de la regresión. Las observaciones influyentes no sólo incluyen potencialmente atípicos y puntos de apalancamiento, sino que pueden incluir otras observaciones. Además, no todos los atípicos y puntos de apalancamiento son necesariamente observaciones influyentes.

Las observaciones influyentes pueden mostrar muchas formas. La Figura 4.8 muestra varias formas de observaciones influyentes y su correspondencia con los residuos. En cada caso, el residuo de los puntos influyentes (la distancia perpendicular desde el punto a la línea de la regresión estimada) no se espera que sea muy grande como para clasificarse como atípico. Por tanto, centrándose sólo en grandes residuos, generalmente ignoraríamos estas observaciones influyentes adicionales. En la Figura 4.8a, el punto de influencia es «bueno», reforzando la pauta general de los datos y reduciendo el error estándar de la predicción y los coeficientes. Es un punto de apalancamiento pero tiene un valor residual pequeño o casi cero, en la medida en que es bien predicho por el modelo de regresión. Sin embargo, los puntos influyentes pueden tener también un efecto que es contrario a la pauta general de los datos restantes y sin embargo tener residuos pequeños (véase Figura 4.8b y 4.8c). En la Figura 4.8b, dos observaciones influyentes son las que cuentan casi por completo para la relación observada, dado que sin ellas no surge una pauta real del resto de los datos. No se identificarían tampoco si sólo se considerasen los residuos grandes, dado que su valor residual sería pequeño. En la Figura 4.8c, se ve un efecto incluso más profundo donde las observaciones influyentes contrarrestan la pauta general del resto de los datos. En este caso, los datos «reales» tendrían mayores residuos que los puntos influyentes «malos». Las observaciones influyentes pueden afectar sólo a una parte de los resultados, como en la Figura 4.8d, donde la pendiente permanece constante pero se desplaza la constante. Finalmente, los puntos influyentes múltiples pueden reforzar el mismo resultado. En la Figura 4.8e, dos puntos influyentes pueden tener la misma relación relativa, haciendo la detección más difícil. Y en la Figura 4.8f, los puntos influyentes tienen posiciones muy diferentes pero un efecto similar los resultados. Esto ejemplos ilustran que debemos desarrollar un instrumental superior de métodos para identificar estos casos influyentes.

Los procedimientos para identificar todo tipo de observaciones influyentes son muy numerosos aunque están peor definidos que muchos otros aspectos del análisis de regresión. Todos los programas informáticos ofrecen un análisis de residuos de los que aquellos con mayores valores (particularmente los residuos estandarizados mayores que 2,0) pueden identificarse fácilmente. Más aún, la mayoría de los programas informáticos ofrecen hoy en día al menos alguna medida de diagnóstico para la identificación de los puntos de apalancamiento y otras observaciones influyentes.

La necesidad de un estudio adicional de los puntos de apalancamiento y los influyentes se pone de manifiesto cuando vemos la sustancial medida en la que la generalización de los resultados y las conclusiones sustantivas (la importancia de las variables, nivel de ajuste, etc.) puede modificarse por un número relativamente pequeño de observaciones. Sean «buenas» (acentuando los resultados) o «malas» (cambiando sustancialmente los resultados), estas observaciones deben identificarse para evaluar su impacto. Influyentes, atípicos y puntos de apalancamiento se basan en alguna de las cuatro condiciones siguientes:

1. Un error en la entrada de observaciones o datos.
2. Una observación válida aunque excepcional que es explicable por una situación extraordinaria.
3. Una observación excepcional sin una explicación plausible.
4. Una observación ordinaria en sus características individuales pero excepcional en su combinación de características.

Pueden recomendarse varios cursos de acción para tratar con las observaciones influyentes de los diferentes tipos. Para un error en la observación se puede corregir el dato o eliminar la observación. Con observaciones válidas pero excepcionales (condición 2), está autorizada la eliminación del caso a menos que las variables que reflejan la situación extraordinaria se incluyan en la ecuación de regresión. La observación inexplicable (condición 3) presenta en un problema espe-

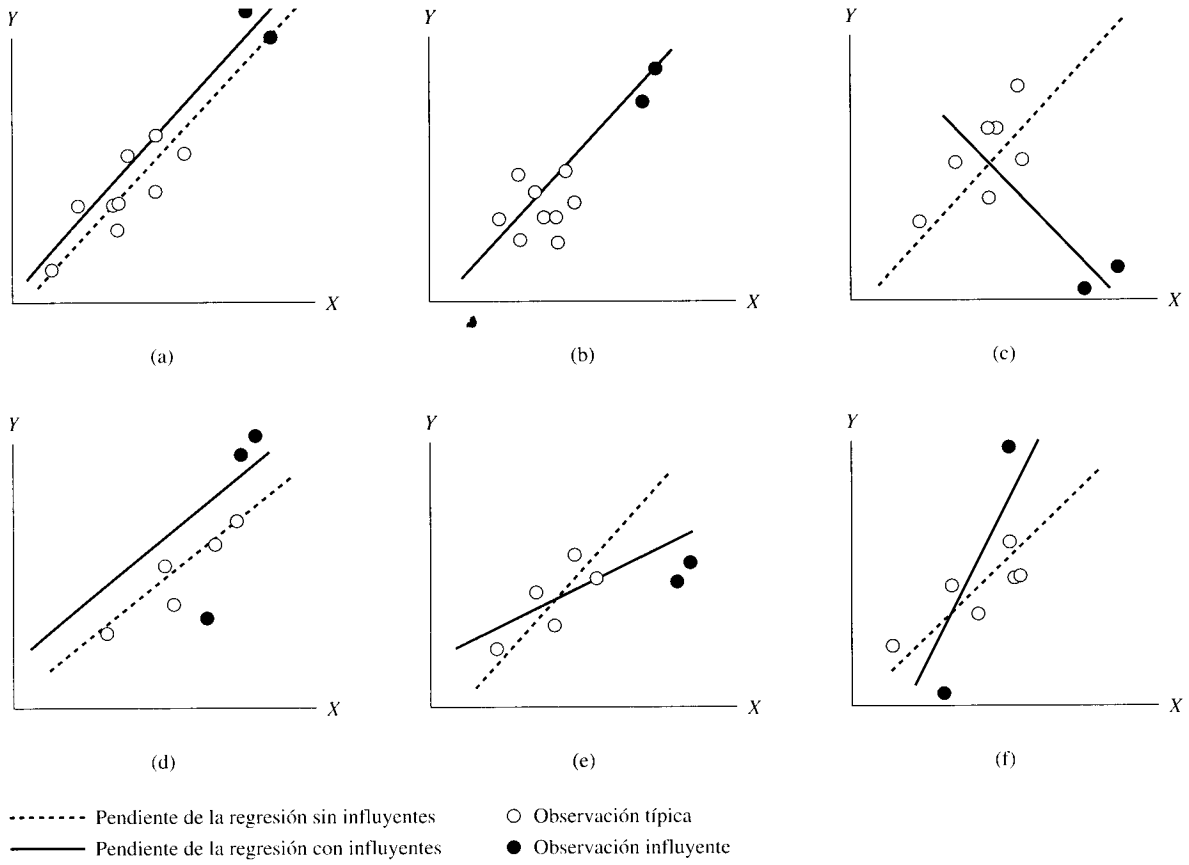


FIGURA 4.8. Formas de las observaciones influyentes.
Adaptado de Beysle *et. al.* y Mason y Perrault [2,9]

cial, dado que no existe razón para eliminar el caso, aunque su inclusión no puede justificarse. Finalmente, la observación que es ordinaria en una variable aunque excepcional en su combinación de características (condición 4) indica modificaciones para la base conceptual del modelo de regresión y debería retenerse.

En todas las situaciones, se recomienda al investigador que elimine las observaciones verdaderamente excepcionales pero que esté en guardia contra la destrucción de observaciones que, aunque diferentes, sean representativas de la población. Recordemos que el objetivo es asegurar el modelo más representativo para la muestra de datos de tal forma que refleje de la mejor forma posible la población de la que se ha extraído. Esto se extiende incluso al mejor ajuste predictivo, dado que ciertos atípicos pueden ser casos válidos que el modelo intentaría predecir, aunque sea pobremente. El investigador debería ser consciente de los casos donde los resultados cambiarían sustancialmente destruyendo sólo una observación aislada o un número muy pequeño de observaciones.

Quinto paso: Interpretación del valor teórico de la regresión

La siguiente tarea del investigador es interpretar el valor teórico de la regresión evaluando los coeficientes de regresión estimados para la explicación de la variable dependiente. Como veremos a lo largo de nuestra exposición, el investigador debe evaluar no sólo el modelo de regresión que se estimó sino también el potencial de variables independientes que se omitieron si se empleó una aproximación combinatorial o de búsqueda secuencial. En esas aproximaciones, la multicolinea-

lidad puede afectar sustancialmente a las variables incluidas en última instancia en el valor teórico de la regresión. Por tanto, además de evaluar los coeficientes estimados, el investigador debe evaluar también el impacto potencial de las variables omitidas para asegurar que la significación práctica se evalúa a la vez que la significación estadística.

Utilización de los coeficientes de regresión

Los coeficientes de regresión estimados se usan para calcular los valores de la predicción para cada observación y para expresar el cambio esperado de la variable dependiente para cada unidad de cambio en las variables independientes. Además de hacer la predicción, nos gustaría saber qué variable independiente es la más útil en la predicción de la variable dependiente. En el ejemplo de la regresión múltiple discutido anteriormente, nos gustaría saber qué variable —el tamaño de la familia o la renta familiar— es más útil en la predicción del número de tarjetas de crédito mantenidas por una familia.

Desafortunadamente, los coeficientes de regresión (b_0, b_1 y b_2) no nos ofrecen esta información. Para ilustrar por qué, podemos utilizar un caso bastante obvio. Supongamos que deseamos predecir los gastos mensuales de los jóvenes en CD (Y), utilizando dos variables independientes; X_1 es la renta de los padres en miles de dólares y X_2 es la asignación mensual de los adolescentes medida en dólares. Supongamos que encontramos el siguiente modelo por un procedimiento de los mínimos cuadrados ordinarios:

$$Y = -0{,}01 + X_1 + 0{,}001X_2$$

Puede asumirse que X_1 es más importante porque su coeficiente es 1.000 veces mayor que el coeficiente de X_2. Este supuesto es cierto, desde luego. Un incremento de 10$ en la renta de los padres produce un cambio en $1 \times 10\$ \div 1.000\$$ en las compras medias de CD (dividimos 10$ por 1.000 porque el valor X_1 se mide en miles de dólares). Un cambio en 10$ en la asignación mensual de los jóvenes produce un cambio de $(0{,}001 \times 10\$)$ en los gastos medios en CD o un cambio de 0,01 en el número medio de CD (porque la asignación de los jóvenes se mide en dólares). Un cambio de 10$ en la renta de los padres produce el mismo efecto de un cambio de 10$ en la asignación de los jóvenes. Ambas variables son igualmente importantes, pero los coeficientes de regresión no revelan directamente este hecho. Podemos resolver este problema mediante el uso de un coeficiente de regresión modificado llamado el *coeficiente beta*.

Estandarización de los coeficientes de regresión: Los coeficientes beta

Si cada una de nuestras variables predictor ha sido **estandarizada** antes de estimar la ecuación de regresión, nos encontraríamos con diferentes coeficientes de regresión. Los coeficientes resultantes de los datos estandarizados se denominan **coeficientes beta**. Su valor reside en que eliminan el problema de tratar con diferentes unidades de medida (como se ha ilustrado previamente), y reflejan el impacto relativo sobre la variable criterio de un cambio en una desviación estándar de cada variable. Ahora que tenemos una unidad común de medida, podemos determinar qué variable es la más influyente.

Se deben tener en cuenta tres precauciones cuando se utilizan los coeficientes beta. En primer lugar, deberían utilizarse como guía de la importancia relativa de las variables individuales independientes únicamente cuando la colinealidad es mínima. En segundo lugar, los valores beta pueden interpretarse sólo en el contexto de las otras variables de la ecuación. Por ejemplo, un valor beta para el tamaño familiar refleja su importancia sólo con relación a la renta familiar, no en ningún sentido absoluto. Si se añade otra variable independiente a la ecuación, el coeficiente beta para el tamaño familiar probablemente cambiaría, dado que podría existir una cierta relación entre el

tamaño de la familia y la nueva variable independiente. La tercera precaución es que los niveles (es decir, familias de 5, 6 y 7) afectan al valor beta. Si encontramos familias de tamaño 8, 9 o 10, el valor beta probablemente cambiará. En resumen, utilizar los coeficientes beta sólo como guía de la importancia relativa de las variables predictor incluidas en tu ecuación, y sólo en el rango de valores para el que realmente existe una muestra de datos.

Evaluación de la multicolinealidad

Un supuesto clave en la interpretación del valor teórico de la regresión es la correlación entre las variables predictor. Se trata de un problema de datos, no un problema de especificación del modelo. La situación ideal para un investigador sería tener una cantidad de variables independientes que estuvieran altamente correlacionadas con la variable dependiente, pero con poca correlación entre sí. Sin embargo, en la mayoría de las situaciones, especialmente las situaciones que incluye datos de respuesta de consumidores, habrá algo de multicolinealidad. En otras ocasiones, como las de la utilización de variables ficticias para representar variables no métricas o términos polinomiales para efectos no lineales, el investigador crea situaciones de alta multicolinealidad. La tarea del investigador es (1) valorar el grado de multicolinealidad y (2) determinar su impacto en los resultados y las soluciones pertinentes en caso de necesidad. En las siguientes secciones abordaremos los efectos de la multicolinealidad y después detallaremos algunos de los procedimientos de diagnóstico útiles y los posibles remedios.

Los efectos de la multicolinealidad

Se puede clasificar los efectos de la multicolinealidad en términos de *explicación* y *estimación.* Los efectos sobre la explicación conciernen principalmente a la capacidad del procedimiento de regresión y a la capacidad del investigador para representar y comprender los efectos de cada variable independiente en el valor teórico de regresión. Conforme ocurre la multicolinealidad (incluso para niveles relativamente bajos de aproximadamente 0,30), el proceso para la separación de los efectos de los individuos es cada vez más difícil. En primer lugar, limita el tamaño del coeficiente de determinación y hace más difícil añadir una predicción explicatoria extra con variables adicionales. En segundo lugar, y al menos tan importante, hace difícil determinar la contribución de cada variable debido a que los efectos de las variables independientes son «mixtos» o se confunden. La multicolinealidad tiene como resultado porciones más grandes de varianza compartida y niveles más bajos de varianza única a partir de los cuales se pueden determinar los efectos de las variables independientes individuales. Por ejemplo, supongamos que una sola variable independiente (X_1) tiene una correlación de 0,60 con la variable dependiente y una segunda variable independiente (X_2) tiene una correlación de 0,50. Entonces X_1 explicaría un 36 por ciento (obtenido con la elevación al cuadrado de la correlación de 0,60) de la varianza de la variable dependiente y X_2 explicaría un 25 por ciento (la correlación de 0,50 elevado al cuadrado). Si las dos variables independientes noestán correlacionadas la una con la otra en absoluto, no existe un «solapamiento» o contribución, de su poder de predicción. La explicación total sería la suma de las dos o el 61 por ciento. Peroconforme aumenta la colinealidad, alguna contribución del poder de predicción tiene lugar, y el poder de predicción colectivo de las variables independientes disminuye. El Apartado 4.1 proporciona más detalle sobre el cálculo de las predicciones de varianza únicas y compartidas entre las variables independientes correlacionadas.

La Figura 4.9 muestra las proporciones de la varianza compartida y única para nuestro ejemplo de dos variables independientes en los casos diversos de colinealidad. Si la colinealidad de estas variables es cero, entonces las variables individuales predicen un 36 y 25 por ciento de la varianza en la variable dependiente, para una predicción global (R^2) de 61 por ciento. Pero conforme aumenta la multicolinealidad, la varianza total explicada disminuye. Además, la cantidad de varianza única para las variables independientes se reduce a niveles que contribuyen al hecho de que los efectos individuales sean bastante problemáticos.

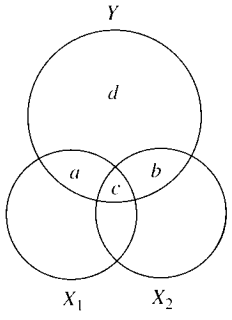
APARTADO 4.1

Cálculo de la varianza única y compartida entre las variables independientes

La base para la estimación de todas las relaciones de regresión es la correlación, que mide la asociación entre dos variables. En el análisis de la regresión, las correlaciones entre las variables independientes y las variables dependientes proporcionan la base para conformar la regresión del valor teórico mediante la estimación de los coeficientes de regresión (ponderaciones) para cada variable independiente que maximice la predicción (varianza explicada) de la variable dependiente. Cuando el valor teórico contiene una única variable independiente, el cálculo de los coeficientes de regresión es directo y se basa en la correlación directa o univariante entre la variable independiente y dependiente. El porcentaje de la varianza explicada de la variable dependiente es simplemente el cuadrado de la correlación directa.

Pero a medida que se añaden variables independientes al valor teórico, los cálculos deben considerar también las intercorrelaciones entre las variables independientes. Si las variables independientes están correlacionadas, entonces «comparten» algo de su poder predictivo. Dado que usamos sólo la predicción del conjunto del valor teórico, la varianza compartida no debe ser «contabilizada dos veces» mediante el uso de las correlaciones directas. Por tanto, calculamos correlaciones adicionales para representar estos efectos compartidos. La primera es **el coeficiente de correlación parcial**, que es la correlación de una variable independiente (X_i) y una variable dependiente (Y) cuando se han suprimido los efectos de las otras variables independientes tanto de X_i e Y. Una segunda forma de correlación es la **correlación semiparcial**, que refleja la correlación entre una variable dependiente e independiente mientras se controlan los efectos predictivos de todas las otras variables independientes sobre X_i. Las dos formas de correlación difieren en que la correlación parcial elimina los efectos de las otras variables independientes sobre X_i e Y, mientras que la correlación semiparcial elimina sólo los efectos sobre X_i. La correlación parcial representa el aumento en el efecto predictivo de una variable independiente respecto al efecto colectivo del resto y se utiliza para identificar variables independientes que tienen el mayor aumento de poder predictivo dado un conjunto de variables independientes ya incluidos en el valor teórico. La correlación semiparcial representa la relación única prevista por una variable independiente una vez que se ha obtenido la predicción compartida con el resto de las variables independientes. Por tanto, la correlación semiparcial se utiliza para el «reparto» de la varianza entre las variables independientes. Elevando al cuadrado la correlación semiparcial obtenemos la varianza única explicada por la variable independiente.

El diagrama adjunto representa la varianza única y compartida entre los variables independientes correlacionadas.



a = varianza de Y explicada únicamente por X_1
b = varianza de Y explicada únicamente por X_2
c = varianza de Y explicada conjuntamente por X_1 y X_2
d = varianza de Y no explicada por X_1 o por X_2

La varianza asociada con la correlación parcial de X_2 controlada por X_1 puede representarse como $b \div (d + b)$, donde $d + b$ representa la varianza sin explicar después de tener en cuenta X_1. La parte de la correlación de X_2 controlada por X_1 es $b \div (a + b + c + d)$, donde $a + b + c + d$ representa el total de la varianza de Y y b en la cantidad únicamente explicada por X_2.

El investigador puede también determinar la varianza única y compartida para las variables independientes a través de cálculos simples. La correlación semiparcial entre la variable dependiente (Y) y una variable independiente (X_1) mientras se controla una segunda variable independiente (X_2) se calcula mediante la siguiente ecuación:

$$\text{Correlación semiparcial de } Y, X_1, \text{ dado } X_2 = \frac{\text{Corr de } Y, X_1 - (\text{Corr de } Y, X_2 * \text{Corr de } X_1X_2)}{\sqrt{1{,}0 - (\text{Corr de } X_1X_2)^2}}$$

Un ejemplo simple con dos variables independientes (X_1 y X_2) ilustrará el cálculo de ambos tipos de varianzas de la variable dependiente (Y). Las correlaciones directas y la correlación entre X_1 y X_2 se muestran en la siguiente matriz de correlación:

	Y	X_1	X_2
Y	1,0		
X_1	0,6	1,0	
X_2	0,5	0,7	1,0

Las correlaciones directas de 0,60 y 0,50 representan relaciones claramente fuertes con Y, mientras que la correlación de 0,70 entre X_1 y X_2 significa que una parte sustancial de esta potencia predictiva puede estar compartida. La correlación semiparcial de X_1 e Y controlando X_2 ($r_{Y, X_1(X_2)}$) y la varianza única prevista por X_1 puede calcularse como:

$$r_{Y, X_1(X_2)} = 0{,}60 - \frac{(0{,}50)(0{,}70)}{\sqrt{1{,}0 - 0{,}70^2}} = 0{,}35$$

Varianza única prevista por $X_1 = 0{,}35^2 = 0{,}1225$

Dado que la correlación directa de X_1 e Y es 0,60, sabemos también que la varianza total prevista por X_1 es $0{,}60^2$ o 0,36. Si la varianza única es 0,1225, entonces la varianza compartida debe ser 0,2375 (0,36 – 0,1225).

Podemos calcular la varianza única explicada por X_2 y confirmar la cantidad de varianza compartida mediante lo siguiente:

$$r_{Y, X_1(X_2)} = \frac{0{,}50 - (0{,}60)(0{,}70)}{\sqrt{1{,}0 - 0{,}70^2}} = 0{,}11$$

Varianza única prevista por $X_1 = 0{,}11^2 = 0{,}0125$

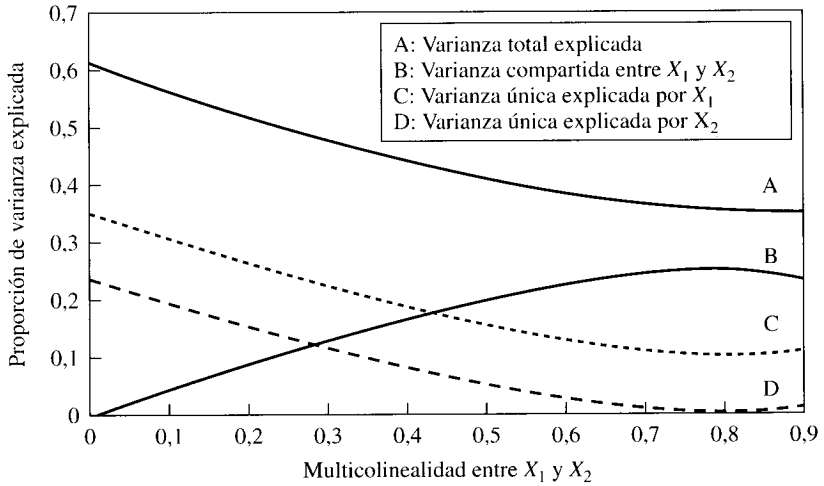
Con la varianza total explicada por X_2 siendo $0{,}50^2$ o 0,25, la varianza compartida calculada como 0,2375 (0,25 – 0,0125). Esto confirma la cantidad encontrada en los cálculos de X_1.

Por tanto, la varianza total explicada (R^2) mediante dos variables independientes es

Varianza única explicada por X_1	0,1225
Varianza única explicada por X_2	0,0125
Varianza compartida explicada por X_1 y X_2	0,2375
Varianza total explicada por X_1 y X_2	0,3725

Estos cálculos pueden extenderse a más de dos variables, pero en la medida en que el número de variables independientes aumenta, es fácil dejar a los programas estadísticos realizar los cálculos.

El cálculo de la varianza única y compartida ilustra el efecto de la multicolinealidad sobre la capacidad de las variables independientes para predecir la variable dependiente. La Figura 3.1 muestra estos efectos cuando nos enfrentamos con niveles de multicolinealidad altos o bajos.



Correlación entre las variables dependientes e independientes:
X_1 y dependiente (0,60), X_2 y dependiente (0,50)

FIGURA 4.9. Proporciones de varianza únicas y compartidas por niveles de la multicolinealidad.

Además de los efectos sobre la explicación, la multicolinealidad puede tener efectos sustantivos sobre la estimación de las coeficientes de regresión y sus pruebas de significación estadísticas. En primer lugar, el caso extremo de la multicolinealidad en la cual dos o más variables están perfectamente correlacionadas, que se domina **singularidad**, impide la estimación de cualquier coeficiente. En este caso, la singularidad tiene que ser eliminada para que la estimación de las coeficientes pueda proceder. Incluso si la multicolinealidad no es perfecta, altos grados de multicolinealidad pueden tener como resultado la incorrecta estimación de los coeficientes de regresión. El siguiente ejemplo (véase Tabla 4.9) ilustra este punto. Con el examen de la matriz de correlación y las regresiones simples es evidente que la relación entre Y y V_1 es positiva, mientras que la relación entre Y y V_2 es negativa. La ecuación de regresión múltiple, sin embargo, no mantiene las relaciones de las regresiones simples. El observador casual que examina solamente las coeficientes de regresión múltiple pensaría que ambas relaciones (Y y V_1, Y y V_2) son negativas, mientras que sabemos que esto no es el caso para Y y V_1. La señal de el coeficiente de regresión de V_1 es equivocada en un sentido intuitivo, pero la fuerte correlación negativa entre V_1 y V_2 tiene como resultado el reverso de las señales para V_2. A pesar de que estos efectos sobre el procedimiento de estimación ocurren principalmente a niveles altos de multicolinealidad (por encima de 0,80) la posibilidad de resultados engañosos exige un escrutinio de cada valor teórico de regresión para una posible multicolinealidad.

TABLA 4.9. Estimaciones de regresión con datos multicolineales

A. DATOS

	Variables en el análisis de regresión		
	Variables dependientes	*Variables independientes*	
Encuestado	Y	V_1	V_2
1	5	6	13
2	3	8	13
3	9	8	11
4	9	10	11
5	13	10	9
6	11	12	9
7	17	12	7
8	15	14	7

B. MATRIZ DE CORRELACIÓN

	Y	V_1	V_2
Y	1,000		
V_1	0,823	1,000	
V_2	−0,977	−0,913	1,000

C. ESTIMACIONES DE REGRESIÓN

Regresión simple (V_1):	$Y = -4{,}75 + 1{,}5V_1$
Regresión simple (V_2):	$Y = -9{,}75 + -1{,}95V_2$
Regresión múltiple (V_1 y V_2):	$Y = 44{,}75 + -0{,}75V_1 + -2{,}7V_2$

La identificación de la multicolinealidad

Hemos visto que los efectos de la multicolinealidad pueden ser sustanciales. En múltiples análisis de regresión, la evaluación de la multicolinealidad se debe realizar en dos pasos: (1) identificación de la magnitud de la colinealidad y (2) la evaluación del grado en que los coeficientes estimados se ven afectados. Si se dicta una acción correctiva, existen varias alternativas. Aquí discutiremos la identificación y los procedimientos de evaluación, y examinaremos entonces algunos remedios posibles.

El medio más simple y obvio de identificar la colinealidad es una examen de la matriz de correlación de las variables independientes. La presencia de una elevada correlación (generalmente de 0,90 en adelante) es la primera indicación de una elevada colinealidad. La ausencia de elevados valores de correlación no asegura una falta de colinealidad. La colinealidad puede deberse a los efectos combinados de dos o más variables independientes.

Dos de las medidas más comunes para evaluar la colinealidad de parejas o de múltiples variables son (1) el valor de **tolerancia** y (2) su inverso —el **factor de inflación de la varianza (VIF)**. Estas medidas nos dan el grado en el que cada variable independiente se explica por otras variables independientes. En términos simples, cada variable independiente se convierte en una variable criterio y se realiza la regresión con el resto de las variables independientes. La tolerancia es la cantidad de variabilidad de las variables independientes seleccionadas no explicadas por el resto de las variables independientes. Por tanto un valor de tolerancia reducido (y elevados valores VIF) denotan una elevada colinealidad. Puede ponerse un umbral de tolerancia en un valor de 0,10 que corresponde a valores VIF por encima de 10. Cada investigador debe determinar el grado de colinealidad que aceptará, en la medida en que los límites por defecto o los recomendados pueden aceptar todavía una colinealidad sustancial. Por ejemplo, el límite sugerido para el valor de tolerancia de 0,10 corresponde a una correlación múltiple de 0,95. Más aún, una correlación múltiple de 0,9 entre una variable independiente y el resto (similar a la regla que aplicamos en la matriz de correlación pareada) resultaría en un valor de tolerancia de 0,19. Por tanto cualquier variable con un valor de tolerancia por debajo de 0,19 (o por encima de un VIF de 5,3) estaría correlacionada en más de 0,90.

Sugerimos que se especifiquen siempre los valores de tolerancia en los programas de regresión, en la medida en que los valores por defecto que excluyen las variables colineales pueden mantener todavía grados muy elevados de colinealidad. Por ejemplo, el valor de tolerancia por defecto en el SPSS para excluir una variables es 0,0001, lo que significa que aunque más de un 99,99 por ciento de la varianza es prevista por el resto de las variables independientes, la variable podría incluirse en la ecuación de regresión. Pueden realizarse estimaciones de los efectos reales de una elevada colinealidad sobre los coeficientes estimados, pero queda fuera del alcance de este texto (véase Neter *et al.* [11]).

Incluso con los diagnósticos que utilizan VIF o valores de tolerancia, no necesitamos saber qué variables están correlacionadas. Un procedimiento desarrollado por Belsley *et. al.* [2] permite la identificación de variables intercorrelaciondas, incluso si tenemos correlaciones entre varias variables. Proporciona al investigador un diagnóstico de mayor potencia en la evaluación de la medida e impacto de la multicolinealidad y se discute en el apéndice de este capítulo.

Remedios para la multicolinealidad

Los remedios para la multicolinealidad se extienden desde una modificación del valor teórico de regresión hasta el uso de los procedimientos de estimación especializada. Una vez que se ha determinado el grado de colinealidad, el investigador tiene varias opciones:

- Omitir una o varias variables independientes correlacionadas e identificar otras variables independientes para ayudar con la predicción. Sin embargo, el investigador debe tener cuidado cuando sigue esta opción, para evitar la creación de un error de especificación cuando se eliminan una o más variables independientes.
- Utilizar el modelo con las variables correlacionados sólo para predecir (es decir, no intentando interpretar los coeficientes de regresión).
- Utilizar las correlaciones simples entre cada variable independiente y cada variable dependiente para entender la relación entre la variable independiente-dependiente.
- Utilizar un método más sofisticado de análisis, como una regresión bayesiana (o un caso especial —regresión de tipo cresta—) o una regresión de componentes principales para obtener un modelo que refleje más claramente los efectos simples de las variables independientes. Estos procedimientos se discuten con más detalle en varios textos [2, 11].

Cada una de estas opciones exige al investigador pronunciarse sobre las variables incluidas en el valor teórico de regresión, que debería ser guiada siempre por el marco teórico del estudio.

Sexto paso: Validación de los resultados

Después de identificar nuestro mejor modelo de regresión, el paso final consiste en asegurarse de que represente a la población general (generalización) y que sea apropiada para situaciones en las cuales será utilizada (transferibilidad). La mejor guía es ver en qué medida se ajusta a un modelo teórico o a un conjunto de resultados validados previamente sobre el mismo asunto. En muchos casos, sin embargo, anteriores resultados o la teoría no están disponibles. Por tanto, discutiremos también las aproximaciones empíricas a la validación del modelo.

Muestras adicionales o muestras divididas

La aproximación más apropiada para la validación empírica es contrastar el modelo de regresión mediante la extracción de una nueva muestra de la población general. Una nueva muestra asegurará la representatividad y puede utilizarse en varias formas. En primer lugar, el modelo original puede predecir valores con la nueva muestra, además del ajuste predictivo. En segundo lugar, se puede estimar un modelo separado con la nueva muestra para a continuación compararla con la ecuación original sobre las características de las variables incluidas: signo, tamaño e importancia relativa de las variables y precisión predictiva. En ambos casos se determina la validez del modelo original comparándolo con los modelos de regresión estimados con la nueva muestra.

Muchas veces la capacidad de recoger nuevos datos está limitada o impedida por factores tales como los costes, la falta de tiempo o la disponibilidad de los encuestados. Cuando este es el caso, se puede entonces dividir la muestra en dos partes: una submuestra de estimación para crear el modelo de regresión y una submuestra de validación/duración utilizada para «contrastar» la ecuación. Existen muchos procedimientos, tanto aleatorios como sistemáticos, para dividir los datos mediante la extracción de dos muestras independientes del conjunto de datos. Todos los programas informáticos de amplia difusión tienen opciones específicas que permiten la estimación y la validación de las submuestras por separado. Véase el Capítulo 5 para una discusión del uso de las submuestras de validación y estimación.

Tanto si se extrae una nueva muestra como si no, es probable que existan diferencias entre el modelo original y otros esfuerzos de validación. El investigador pasa ahora a ser un mediador entre los distintos resultados, buscando el mejor modelo en las diferentes muestras. La necesidad de continuos esfuerzos de validación y de refinamiento del modelo nos recuerda que ningún modelo de regresión, a menos que se estime del conjunto de la población, es el modelo final y absoluto.

Cálculo del estadístico PRESS

Una aproximación alternativa a la obtención de muestras adicionales con el fin de validar el modelo es emplear la muestra original de forma especializada mediante el cálculo del **estadístico PRESS**, una medida similar a R^2 utilizada para evaluar la precisión predictiva del modelo de regresión estimado. Difiere de las principales aproximaciones en que se estima no uno, sino $n - 1$ modelos de regresión. El procedimiento, similar a las técnicas de «bootstrapping» tratadas en el Capítulo 12, omite una observación en la estimación del modelo de regresión y predice a continuación las observaciones omitidas con el modelo estimado. De esta forma, la observación no puede afectar a los coeficientes del modelo utilizado para calcular su valor predictivo. El procedimiento se aplica otra vez, omitiendo otra observación, estimando un nuevo modelo y realizando la predicción. Pueden sumarse los residuos de las observaciones para obtener una medida conjunta del ajuste predictivo.

Comparación de los modelos de regresión

Cuando comparamos modelos de regresión, el término de comparación más común suele ser el ajuste predictivo conjunto. Discutimos previamente que el coeficiente de determinación (R^2) nos ofrece esta información, pero tiene un inconveniente: por muchas variables que se incluyan, nunca puede disminuir. Por tanto, aunque incluyéramos todas las variables independientes, nunca encontraríamos un R^2 más elevado, aunque podamos conseguir el mismo R^2 con un número más pequeño de variables o encontrar que un número más pequeño de predictores arroja casi el mismo valor. Por tanto, para comparar entre modelos con diferentes números de predictores, utilizaremos el R^2 ajustado, que es útil en la comparación de modelos entre diferentes conjuntos de datos, en la medida en que se compensarán los diferentes tamaños muestrales.

Predicción con el modelo

Las predicciones del modelo siempre pueden realizarse aplicando el modelo estimado para un nuevo conjunto de valores de las variables independientes y calculando los valores de la variables criterio. Sin embargo, al hacerlo debemos considerar varios factores que pueden tener un serio impacto en la calidad de las nuevas predicciones:

1. Cuando aplicamos el modelo a una nueva muestra, debemos recordar que las predicciones contienen ahora no sólo las variaciones muestrales respecto de la muestra original sino también la muestra nuevamente extraída. Por tanto deberíamos calcular siempre los intervalos de confianza de nuestras predicciones además de los puntos estimados para ver el rango esperado de los valores criterio.
2. Tenemos que asegurarnos de que las condiciones y relaciones medidas en el momento en que la muestra original fue tomada no han cambiado. Por ejemplo, en nuestro ejemplo de las tarjetas de crédito, si la mayor parte de las compañías comenzaran a cargar mayores comisiones por las tarjetas de crédito, la posesión efectiva de tarjetas de crédito puede cambiar sustancialmente, aunque esta información no esté incluida en el modelo.
3. Finalmente, no deberíamos utilizar el modelo para estimar más allá del rango de las variables independientes que se encuentran en la muestra. Por ejemplo, si las familias más grandes tienen seis miembros, puede ser imprudente predecir la posesión de tarjetas de crédito para familias de diez miembros. Uno no puede suponer que las relaciones son las mismas para valores de las variables independientes sustancialmente superiores que aquellos de la muestra original.

Ejemplo de un análisis de regresión

Se han discutido en las secciones precedentes los temas que conciernen a la aplicación e interpretación del análisis de regresión siguiendo una modelización construida en seis pasos, que fue introducida en el Capítulo 1. Para ilustrar las cuestiones importantes de cada paso, en las secciones siguientes utilizaremos un ejemplo con el que detallaremos la aplicación de la regresión multivariante a los problemas de investigación identificados por HATCO. El Capítulo 1 introduce un marco de investigación en el que HATCO ha obtenido unas medidas en un estudio sobre clientes. Para demostrar el uso de la regresión múltiple, veremos los procedimientos utilizados por el investigador para intentar la predicción de los niveles de fidelidad al producto de los individuos de la muestra con un conjunto de siete observaciones independientes.

Primer paso: Objetivos de la regresión múltiple

La gestión de HATCO se ha interesado hace tiempo en la precisión predictiva del nivel de actividad obtenida a partir de la información suministrada por sus clientes en un intento de proporcio-

nar un mejor fundamento de los controles de producción y el marketing. Con este fin, los investigadores de HATCO proponen que se debería intentar la realización de un análisis de regresión múltiple para predecir los niveles de fidelidad a los productos por parte de los clientes basándose en las percepciones que tienen éstos de la actividad de HATCO. Además de buscar una predicción precisa de los niveles de uso, los investigadores también estaban interesados en identificar los factores que llevan al aumento de la utilización del producto para su aplicación en campañas de marketing diferenciadas.

Para aplicar el procedimiento de regresión, el investigador selecciona la fidelidad al producto (X_9) como la variable dependiente (Y) a predecir por las variables independientes representadas por las percepciones de la actividad de HATCO. Las siguientes siete variables se incluyeron como variables independientes:

X_1 Velocidad de entrega
X_2 Nivel de precios
X_3 Flexibilidad de precios
X_4 Imagen del fabricante
X_5 Servicio
X_6 Imagen de la fuerza de ventas
X_7 Calidad del producto

La relación entre las siete variables predictor y la fidelidad al producto se asumía que era estadística y no funcional, dado que suponen percepciones acerca de la actividad que incluyen errores del nivel de medida.

Segundo paso: Diseño de la investigación mediante el análisis de regresión múltiple

El estudio de HATCO obtuvo 100 encuestas del total de su base de clientes. Los 100 encuestados proporcionaron respuestas completas, resultando 100 observaciones disponibles para el análisis. La primera cuestión a responder en relación con el tamaño muestral es el nivel de relación (R^2) que puede ser razonablemente detectado con el análisis de regresión propuesto. La Tabla 4.7 nos indica que la muestra de 100, con siete variables potencialmente independientes, es capaz de detectar relaciones con valores de R^2 aproximadamente un 30 por ciento para una potencia de 0,80 con un nivel de significación fijado en 0,01. Si el nivel de significación se relaja a 0,05, entonces el análisis identificará relaciones que expliquen cerca del 13 por ciento de la varianza. El análisis de regresión propuesto se consideró suficiente para identificar no sólo relaciones estadísticamente significativas sino también relaciones que tuviesen relevancia práctica.

La muestra de 100 observaciones también cumple la pauta propuesta de la ratio de observaciones de las variables independientes con una ratio de 15 a 1. Mientras el investigador puede asegurarse de que no se verá perjudicado por un sobreajuste de la muestra, todavía deberá validar los resultados para asegurar su generalización al conjunto de la base de clientes.

Tercer paso: Supuestos del análisis de regresión múltiple

Es esencial cumplir los supuestos del análisis de regresión para asegurar tanto que los resultados obtenidos sean verdaderamente representativos de la muestra como que hemos obtenido los mejores resultados posibles. Cualquier incumplimiento serio de los supuestos debe detectarse y corregirse si es posible. El análisis que asegura que la investigación cumple los supuestos básicos del análisis de regresión implica dos etapas: contrastación de las variables dependientes e independientes y la contrastación de la relación conjunta después de la estimación del modelo. Esta sección está dedicada a la evaluación de las variables individuales, mientras que el examen de la relación conjunta tendrá lugar después de haber estimado el modelo.

Los tres supuestos imputados a las variables individuales son linealidad, constancia de la varianza y normalidad. A efectos del análisis de regresión, resumiremos los resultados obtenidos en el Capítulo 2 detallando el examen de las variables dependientes e independientes. En primer lugar, los gráficos de dispersión de las variables individuales no indican la existencia de relaciones no lineales entre las variables dependientes y las independientes. Los test de heterocedasticidad encontraron que sólo una de las variables (X_2) incumple este supuesto. Finalmente, en los test de normalidad, se encontró que tres de las variables (X_2, X_4 y X_6) incumplían los test estadísticos. En cada caso, se indicaron las transformaciones logarítmicas. Las series de test de estos tres supuestos subyacentes a los análisis de regresión indicaban que había que centrarse en la normalidad de las tres variables independientes. Aunque el análisis de regresión se ha mostrado bastante robusto, incluso cuando se incumple el supuesto de normalidad, los investigadores estimarían el análisis de regresión tanto con las variables originales como con las transformadas para evaluar las consecuencias de la no-normalidad de las variables independientes sobre la interpretación de los resultados. Haremos una estimación de la ecuación de regresión tanto con las variables originales como con las variables transformadas en la siguiente sección.

Cuarto paso: Estimación del modelo de regresión y valoración global del ajuste

Con el análisis de regresión especificado en términos de variables dependientes e independientes, la muestra considerada adecuada en función de los objetivos del estudio y los supuestos evaluados para las variables individuales, el proceso procede ahora a la estimación del modelo de regresión y la evaluación del ajuste total del modelo. A efectos expositivos, se emplea el método de regresión por etapas para seleccionar las variables a incluir en el valor teórico de la regresión. Después de que el modelo de regresión se ha estimado, el valor teórico será evaluado para ver el cumplimiento de los supuestos del análisis de regresión. Finalmente, se examinarán las observaciones para determinar si cualquier observación debería considerarse influyente. Cada uno de estos supuestos se discutirá en las secciones siguientes.

La estimación por etapas: Selección de la primera variable

La Tabla 4.10 muestra todas las correlaciones entre las siete variables independientes y sus correlaciones con la variable dependiente (Y). El examen de la matriz de correlación indica que el predictor 5 (X_5) está más correlacionado con la variable dependiente (0,70). Nuestro primer paso es construir un modelo de regresión utilizando este mejor predictor. Obsérvese que la correlación del predictor 1 con la variable dependiente es 0,68. Sin embargo, X_1 está correlacionado (0,61) con X_5.

TABLA 4.10. Matriz de correlación: Datos HATCO

	Predictores						
Variables	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Predictores							
X_1 Velocidad de entrega	1,00						
X_2 Nivel de precios	−0,35	1,00					
X_3 Flexibilidad de precios	0,51	−0,49	1,00				
X_4 Imagen del fabricante	0,05	0,27	−0,12	1,00			
X_5 Servicio global	0,61	0,51	0,07	0,30	1,00		
X_6 Imagen del personal de ventas	0,08	0,19	−0,03	0,79	0,24	1,00	
X_7 Calidad de producto	−0,48	0,47	−0,45	0,20	−0,06	0,18	1,00
Dependiente							
$Y(X_9)$ Nivel de uso	0,68	0,08	0,56	0,22	0,70	0,26	−0,19

TABLA 4.11. Resultados del ejemplo: Primer paso del ejemplo de regresión múltiple de HATCO

Variable introducida: X_5 servicio global	
R Múltiple	0,701
R^2 Múltiple	0,491
R^2 Ajustado	0,486
Error estándar de la estimación	6,446

Análisis de la varianza

	Suma de los cuadrados	*gl*	*Cuadrado media*	*Ratio F*
Regresión	3.927,31	1	3.927,31	94,53
Residuos	4.071,69	98	41,55	

	Variables en la ecuación				*Variables no presentes en la ecuación*	
Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente (beta) de regresión estandarizado*	*Valor t parcial*	*Correlación parcial*	*Valor t*
Constante	21,653	2,596		8,341		
X_5 Servicio global	8,384	0,862	0,701	9,722		
X_1 Velocidad de entrega					0,439	4,812
X_2 Nivel de precios					−0,453	−5,007
X_3 Flexibilidad de precio					0,720	10,210
X_4 Imagen del fabricante					0,022	0,216
X_6 Imagen del personal de ventas					0,126	1,252
X_7 Calidad de producto					−0,216	−2,178

Esta es nuestra primera pista de que el uso conjunto de los dos predictores (X_1 y X_5) puede no ser apropiado dado que están altamente correlacionados entre sí tanto como lo están con la variable dependiente. Los resultados de este primer paso aparecen en la Tabla 4.11. Los conceptos de la Tabla 4.11 son los siguientes.

R múltiple R múltiple es el coeficiente de correlación (en este paso) de la regresión simple de X_5 y la variable dependiente. No tiene signo más o menos porque en el análisis múltiple los signos de las variables individuales pueden variar, de tal forma que los coeficientes reflejan el grado de asociación.

R cuadrado El R cuadrado (R^2) es el coeficiente de correlación al cuadrado ($0{,}701^2 \times 0{,}491$), también denominado como coeficiente de determinación. Este valor indica el porcentaje de la variación total de Y explicado por X_5. La suma total de los cuadrados (3.927,31 + 4.071,69 = 7.999,0) es el error al cuadrado que se obtendría si utilizamos sólo la media de Y para predecir la variable criterio. Utilizando los valores de X_5 reduce este error en un 49,1 por ciento (3.927,31 ÷ 7.999,0 = 0,491).

Error estándar de la estimación El error estándar de la estimación es otra medida de la precisión de nuestras predicciones. Es la raíz cuadrada de la suma de los errores al cuadrado dividido por los grados de libertad ($\sqrt{4{,}071 \div 98}$). Representa una estimación de la desviación estándar de los valores dependientes efectivos alrededor de la línea de regresión; esto es, una medida de la variación alrededor de la línea de regresión. El error estándar de la estimación puede también ser

visto como la desviación estándar de los errores de predicción y se convierte en una medida de evaluación del tamaño absoluto del error de predicción. Se utiliza también en la estimación del tamaño del intervalo de confianza para las predicciones. Véase Neter *et al.* [11] para más detalles en relación con este procedimiento.

Variables en la ecuación (primer paso) En el primer paso, una única variable predictor (X_5) se utiliza para calcular la ecuación de regresión con el fin de predecir la variable dependiente. Para cada variable de la ecuación, se requiere definir varias medidas: el coeficiente de regresión, el coeficiente del error estándar y los valores t de las variables de la ecuación.

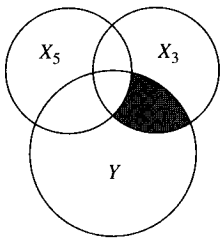
El coeficiente de regresión El valor 8,38 es el coeficiente de regresión (b_5) de la variable predictor (X_5). Por tanto el valor de la predicción para cada variable X_5 es la constante más el valor de la variable predictor por el coeficiente de regresión ($21{,}65 + 8{,}38X_5$). El coeficiente de regresión estandarizado, o valor beta, de 0,70, es el valor calculado a partir de los datos estandarizados. Con sólo una variable independiente, el coeficiente beta al cuadrado iguala al coeficiente de determinación. El valor beta nos permite comparar el efecto de X_5 sobre Y así como el efecto sobre Y del resto de las variables predictor en cada paso, dado que este valor reduce el coeficiente de regresión a una unidad comparable, el número de desviaciones estándar. (Nótese que hasta ahora no tenemos otras variables para comparar)

Error estándar del coeficiente El error estándar del coeficiente es el error estándar de la estimación de b_5. El valor de b_5 dividido por el error estándar ($8{,}38 \div 0{,}86 = 9{,}74$) es el valor t calculado en un test t de hipótesis $b_5 = 0$. Un error estándar más pequeño implica una predicción más segura. Por tanto, preferiríamos tener errores estándar pequeños y menores intervalos de confianza. Este coeficiente es denominado también como el error estándar del coeficiente de regresión; es una estimación de la medida en que el coeficiente variará entre muestras del mismo tamaño obtenidas de la misma población; esto es, si se fueran a tomar múltiples muestras del mismo tamaño de la misma población y utilizarlas para calcular la ecuación de regresión, constituiría una estimación de cómo cambiaría el coeficiente de muestra a muestra.

Valores *t* de las variables de la ecuación El valor t de las variables de la ecuación mide la significación de la correlación parcial de la variable reflejada en el coeficiente de regresión. Es particularmente útil en el quinto paso de la Figura 4.7 en la decisión sobre si una variable debería salir de la ecuación una vez que se ha añadido. También se muestra en la tabla el nivel de significación, que se compara con el umbral fijado por el investigador para eliminar la variable. En nuestro ejemplo, hemos fijado un nivel en 0,10 para eliminar las variables de la ecuación. El valor crítico para un nivel de significación de 0,10 con 98 grados de libertad es 1,658. Por tanto, X_5 cumple nuestros requisitos para incluirla en la ecuación de regresión. En este punto se dan a menudo los valores F en lugar de los valores t. Son directamente comparables, en la medida en que el valor t es la raíz cuadrada del valor F.

Variables no presentes en la ecuación Aunque se ha incluido X_5 en la ecuación de regresión, quedan otras seis variables potencialmente independientes que se pueden incluir para mejorar la predicción de la variable dependiente. Para esos valores, tenemos dos medidas para evaluar su contribución potencial: correlaciones parciales y valores t.

Correlación parcial La correlación parcial es una medida de la variación de Y que no se tiene en cuenta por las variables de la ecuación (sólo X_5 en el primer paso) que pueden explicarse por cada una de estas variables adicionales. Por ejemplo, el valor 0,720 representa la correlación parcial de X_3, dado que X_5 está en la ecuación. Recordemos que la correlación parcial puede ser mal interpretada. No quiere decir que explicamos un 72,0 por ciento de la varianza previamente no explicada. Significa que el 51,8 por ciento ($0{,}720^2 = 0{,}518$ el coeficiente de correlación parcial) de la varianza no explicada (no el total) puede explicarse mediante X_3. Dado que el 49,1 por ciento ya se explicó mediante X_5, 26,4 por ciento ($1 - 0{,}491) \times 0{,}518$ del total de la varianza podría explicarse añadiendo la variable X_3. Un diagrama de Venn ilustra perfectamente este concepto:



El área sombreada de X_3 como una proporción del área sombreada de Y representa la correlación parcial de X_3 con Y dado X_5. Esto representa la varianza en X_3 (después de eliminar los efectos de X_5 sobre X_3) en común con la varianza restante en Y (después de eliminar los efectos de X_5 sobre Y). El cálculo de la varianza única asociada con la entrada de X_3 puede determinarse a través de la correlación parcial, como se describe en el Documento 4.1.

Valores t de variables no presentes en la ecuación La columna de valores *t* mide la significación de las correlaciones parciales de las variables no presentes en la ecuación. Éstas se calculan como una ratio de la suma de los cuadrados adicionales explicados por la inclusión de una variable particular y la suma de los cuadrados que queda después de añadir la misma variable. Si este valor *t* no supera un nivel de significación especificado, no se permitirá a la variable entrar en la ecuación. Los valores t de la tabla para un nivel de significación de 0,05 con 97 grados de libertad es de 1,98. Mirando en la columna de valores *t* en la Tabla 4.12, nos damos cuenta que cuatro variables (X_1, X_2, X_3 y X_7) superan este valor y resultan ser candidatas a la inclusión.

TABLA 4.12. Resultados del ejemplo: Segundo paso del ejemplo de regresión múltiple de HATCO

Variable introducida: X_3 flexibilidad de precios	
R Múltiple	0,869
R^2 Múltiple	0,755
R^2 Ajustado	0,750
Error estándar de la estimación	4,498

Análisis de la varianza

	Suma de los cuadrados	*gl*	*Cuadrado media*	*Ratio F*
Regresión	6.036,5	2	3.018,26	149,18
Residuos	1.962,5	97	20,23	

	Variables en la ecuación				*Variables no presentes en la ecuación*	
Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente (beta) de regresión estandarizado*	*Valor t parcial*	*Correlación parcial*	*Valor t*
Constante	−3,489	3,057				
X_3 Flexibilidad de precio	3,336	0,327	0,515	10,210		
X_5 Servicio global	7,974	0,603	0,666	13,221		
X_1 Velocidad de entrega					0,021	0,205
X_2 Nivel de precios					−0,027	−0,267
X_4 Imagen del fabricante					0,181	1,808
X_6 Imagen del personal de ventas					0,236	2,378
X_7 Calidad de producto					0,169	1,683

La Tabla 4.11 muestra que la correlación simple de X_1 con la variable independiente era de 0,68, pero era sólo de 0,56 para X_3. Por tanto, se podía haber pensado que la variable X_1 estaría incluida en el siguiente modelo. Pero al decidir qué variables adicionales se van incluir en la ecuación, seleccionaríamos en primer lugar la variable independiente que muestra la mayor correlación parcial con la variable dependiente (no la mayor correlación con Y). La correlación parcial de X_3 es el mayor (0,720) y por tanto se considerará la inclusión de X_3 (e incluso X_2) al modelo antes que X_1.

Sabemos ahora que una parte significativa de la varianza de la variable dependiente se explica mediante X_5. Podemos también ver que X_3 tiene el mayor coeficiente de correlación parcial con la variable dependiente, y que el valor t es significativo en el nivel 0,05. (Es significativo también en el nivel 0,01). Podemos ahora observar el nuevo modelo utilizando tanto X_5 como X_3.

Estimación por etapas: Adición de X_3

Los valores del R múltiple y del R cuadrado han aumentado con la inclusión de X_3 (véase Tabla 4.12). El R^2 ha aumentado en un 26,4 por ciento, la cantidad que predecimos cuando examinamos el coeficiente de correlación parcial de X_3, que es de 0,720. El aumento en el R^2 de un 26,4 por ciento se deriva de multiplicar el 50,9 por ciento de la variación que no se explicó después del primer paso por el cuadrado de la correlación parcial: $50,9 \times (0,720)^2 = 26,4$; esto es, de un 50,9 por ciento que no se explicaba con X_5, $(0,720)^2$ de esta varianza se explicaba añadiendo X_3, arrojando un rendimiento sobre la varianza total explicada de 0,755 —esto es, $0,491 + [0,509 \times (0,720)^2]$.

El valor de b_5 ha cambiado muy poco. Esto es una indicación de que las variables X_5 y X_3 son relativamente independientes (la correlación simple entre las dos variables es 0,07). Si el efecto de X_3 sobre Y fuera totalmente independiente del efecto sobre X_5, el coeficiente b_5 no cambiaría nada. Los valores de la t parcial indican que tanto X_5 como X_3 son predictores estadísticamente significativos de Y. El valor t de X_5 es ahora 13,221, siendo 9,772 en el primer paso. El valor t para X_3 examina la contribución de esta variable teniendo en cuenta que X_5 ya está en la ecuación. Nótese que el valor t para X_3 (10,210) es el mismo valor mostrado para X_3 en el primer paso bajo el título «variables no incluidas en la ecuación» (véase Tabla 4.11).

Dado que X_3 y X_5 tienen contribuciones significativas, ninguna la explicación de la variación de la variable independiente, podemos preguntarnos si existen otros predictores. Observando las correlaciones parciales de las variables no presentes en la ecuación de la Tabla 4.12, vemos que X_6 tiene la mayor correlación parcial (0,236). Esta variable explicaría el 5,6 por ciento de la hasta ahora no explicada varianza ($0,236^2 = 0,056$) o 1,4 por ciento de la varianza total $[(1 - 0,755) \times 0,56 = 0,14]$. Se trata de una contribución muy modesta al poder explicativo de nuestra predicción, incluso aunque la correlación parcial sea significativa a un nivel de significación del 0,05. (Nota: el valor t de la tabla para 96 grados de libertad para un nivel de 0,05 es 1,98 mientras que el valor t para X_6 es 2,378).

Estimación por etapas: Añadimos la tercera variable X_6

Con la inclusión de X_6 en la ecuación de regresión se muestran los resultados en la Tabla 4.13. El valor de R^2 aumenta en un 1,4 por ciento. Además, el examen de las correlaciones parciales de X_1, X_2, X_4 y X_7 indica que no se obtendrán ganancias adicionales introduciéndolas en la ecuación de predicción. Estas correlaciones parciales son muy pequeñas y tienen unos valores parciales de t asociados a ellos que no son estadísticamente significativos para el nivel de (0,05) elegido para este modelo.

Evaluación del valor teórico respecto de los supuestos del análisis regresivo

Al evaluar la ecuación estimada, hemos considerado la significación estadística. Debemos también dirigirnos a otros dos asuntos elementales: (1) cumplimiento de los supuestos subyacentes en la re-

TABLA 4.13. Resultados del ejemplo: Tercer paso del ejemplo de regresión múltiple de HATCO

Variable introducida: X_6 imagen del personal de ventas

R Múltiple	0,877
R^2 Múltiple	0,768
R^2 Ajustado	0,761
Error estándar de la estimación	4,394

Análisis de la varianza

	Suma de los cuadrados	*gl*	*Cuadrado media*	*Ratio F*
Regresión	6.145,7	3	2.048,6	106,12
Residuos	1.853,3	96	19,3	

	Variables en la ecuación				*Variables no presentes en la ecuación*	
Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente (beta) de regresión estandarizado*	*Valor t parcial*	*Correlación parcial*	*Valor t*
Constante	−6,520					
X_3 Flexibilidad de precio	3,376	0,320	0,521	10,562		
X_5 Servicio global	7,621	0,607	0,637	12,547		
X_6 Imagen del personal de ventas	1,406	0,591	0,121	2,378		
X_1 Velocidad de entrega					0,040	0,389
X_2 Nivel de precios					−0,041	−0,405
X_4 Imagen del fabricante					−0,002	−0,021
X_7 Calidad de producto					0,130	1,273

gresión y (2) identificación de los datos influyentes. Consideraremos cada uno de estos asuntos en las siguientes secciones.

Los supuestos a examinar son linealidad, homocedasticidad, independencia de los residuos y normalidad. La principal medida utilizada en la evaluación del valor teórico de la regresión es el residuo —la diferencia entre la variable dependiente efectiva y su valor predictivo. A efectos comparativos, utilizaremos el residuo basado en la t de Student. En la Figura 4.10 se muestra el tipo más básico de gráfico de residuos. Como podemos ver, los residuos entran dentro de una pauta generalmente aleatoria, muy similar al gráfico de no-correlación de residuos de la Figura 4.5a. Sin embargo, debemos realizar test específicos para cada supuesto para comprobar los posibles incumplimientos.

Linealidad El primer supuesto, la linealidad, se evaluaría a través de los gráficos de regresión parcial y de residuos. La Figura 4.10 no muestra ninguna pauta no lineal entre los residuos, lo que asegura que la ecuación considerada es lineal. Pero también debemos asegurarnos, al utilizar más de una variable predictor, que cada relación de las variables predictor es lineal, así como asegurar su mejor representación en la ecuación. Para hacerlo, utilizamos el gráfico de regresión parcial para cada variable independiente de la ecuación. En la Figura 4.11 vemos que las relaciones de X_3 y X_5 están muy bien definidas; por tanto, tenemos efectos fuertes y significativos en la ecuación de regresión. La variable X_6 está menos definida, tanto en la pendiente como en la distribución de los puntos, lo que explica su menor efecto en la ecuación (evidenciado por el menor coeficiente, valor beta y nivel de significación). Para todas las variables, no se observa una pauta no lineal, lo que supone un cumplimiento del supuesto de linealidad para cada variable predictor.

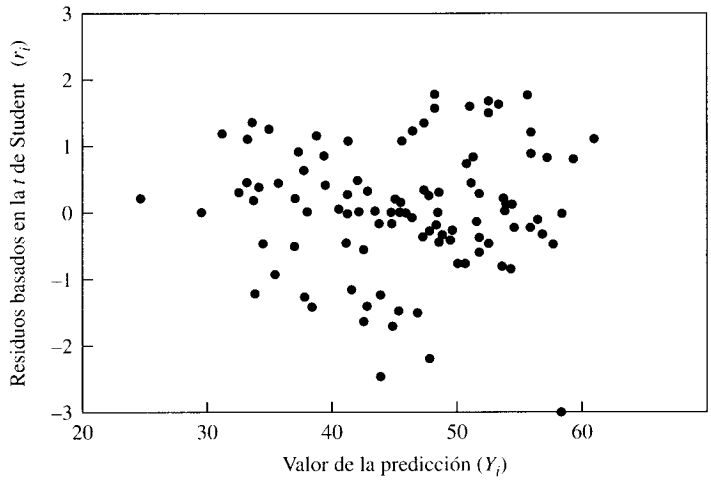


FIGURA 4.10. Análisis basado en la *t* de Student de los residuos.

Homocedasticidad El siguiente supuesto hace referencia a la constancia de los residuos para los valores que van tomando las variables independientes. Nuestro análisis se lleva a cabo de nuevo mediante el examen de los residuos (Figura 4.10), que no muestra una pauta de aumento o disminución de los residuos. Este hallazgo indica homocedasticidad en el caso multivariante (el conjunto de las variables independientes).

Independencia de los residuos El tercer supuesto hace referencia a los efectos de la inercia de una observación a otra, lo que hace que los residuos no sean independientes. Cuando se encuentra este efecto inercia en casos tales como las series temporales, el investigador debe identificar las potenciales variables secuenciales (tales como el tiempo en un problema de series temporales) y el gráfico de residuos de la variable en cuestión. Por ejemplo, supongamos que el número de identificación representa el orden en que recogemos las respuestas. Podemos realizar el gráfico de los residuos y observar si surge algún modelo o pauta. En nuestro ejemplo, se intentó con varias variables, incluyendo el número de identificación y cada variable independiente, y no se encontró una pauta consistente. Recordemos que estamos utilizando los residuos en este análisis y no los valores de las variables dependientes originales, dado que nos centramos en los errores de predicción, no en la relación obtenida en la ecuación de la regresión.

Normalidad El último supuesto que vamos a contrastar es la normalidad del término de error del valor teórico con un examen visual de los gráficos de probabilidad normal de los residuos. Como se muestra en la Figura 4.12, los valores se sitúan a lo largo de la diagonal sin alejamientos sustan-

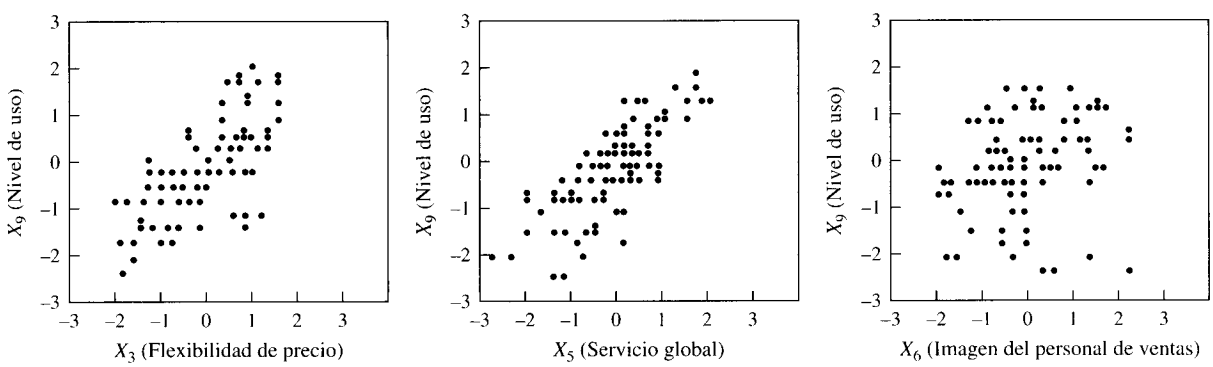


FIGURA 4.11. Gráficos de regresión parcial estandarizada.

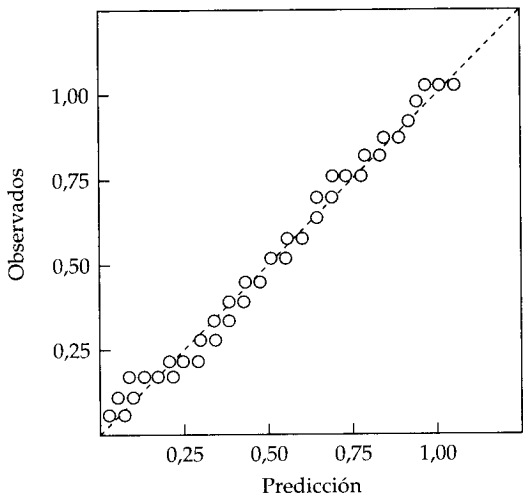


FIGURA 4.12. Gráfico de probabilidad normal: Residuos estandarizados.

ciales o sistemáticos; por tanto, se considera que los residuos representan una distribución normal. Así pues, encontramos que el valor teórico de la regresión cumple el supuesto de normalidad.

Soluciones para el incumplimiento de los supuestos Después de contrastar los incumplimientos de los cuatro supuestos básicos de las regresiones multivariantes tanto para las variables individuales como para el valor teórico, el investigador debería evaluar el impacto de cualquier solución sobre los resultados. Con el examen de las variables individuales en el Capítulo 2, los únicos remedios necesarios fueron la transformación de X_2, X_4 y X_6, con el fin de conseguir la normalidad. Si sustituimos esas variables por sus valores originales y reestimamos la ecuación de la regresión, conseguiremos resultados casi idénticos (véase Tabla 4.14). Entran las mismas variables en la ecuación, siendo la única diferencia sustantiva el ligeramente más potente coeficiente para la transformación de la variable X_6 y una ligera mejoría en el valor de R^2 (0,771 contra 0,768). Las variables independientes no presentes en la ecuación mostrarán todavía niveles no significativos al entrar en la ecuación, incluso aquellas que fueron transformadas. Por tanto, en este caso, las soluciones al incumplimiento de los supuestos mejoran ligeramente la predicción pero no alteran los hallazgos sustantivos.

Identificación de los casos atípicos como observaciones influyentes

En nuestro análisis final, intentaremos identificar cualquier observación que sea influyente (al tener un impacto desproporcionado sobre los resultados de la regresión) y determinar si deberían excluirse del análisis. Aunque existen procedimientos más detallados para la identificación de atípicos como observaciones influyentes, nos centraremos en la siguiente sección en el uso de los residuos en la identificación de atípicos.

Los métodos de diagnóstico más elementales son los residuos y la identificación de cualquier caso atípico —esto es, observaciones no previstas bien por la ecuación de regresión que tienen un gran residuo. La Figura 4.13 muestra los residuos basados en la *t* de Student para cada observación. Dado que los valores corresponden a valores *t*, pueden fijarse los límites superior e inferior en el intervalo de confianza deseado. Quizá el nivel más utilizado sea la confianza del 95 por ciento ($\alpha = 0,05$). El valor correspondiente de *t* es 1,96, identificando por tanto los residuos estadísticamente significativos como aquéllos con residuos mayores que este valor. Cuatro observaciones (7, 11, 14, 100) tienen residuos significativos y pueden clasificarse como casos atípicos. Los casos atípicos son importantes porque son observaciones no representadas por la ecuación de regresión por una o varias razones —una de las cuales puede ser una influencia sobre la ecuación que requiera una solución.

TABLA 4.14. Resultado del ejemplo: Resultados de la regresión múltiple después de aplicar las soluciones correspondientes al incumplimiento de los supuestos

R Múltiple	0,878
R^2 Múltiple	0,771
R^2 Ajustado	0,764
Error estándar de la estimación	4,368

Análisis de la varianza

	Suma de los cuadrados	*gl*	*Cuadrado media*	*Ratio F*
Regresión	6.167,1	3	2.055,71	107,73
Residuos	1.831,9	96	19,08	

	Variables en la ecuación				*Variables no presentes en la ecuación*	
Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente (beta) de regresión estandarizado*	*Valor t parcial*	*Correlación parcial*	*Valor t*
Constante	−6,792	3,226		−2,11		
X_3 Flexibilidad de precio	3,409	0,319	0,526	10,70		
X_5 Servicio global	7,640	0,599	0,639	12,75		
log X_6 Imagen del personal de ventas	3,953	1,511	0,131	2,62		
X_1 Velocidad de entrega					0,048	0,469
log X_2 Nivel de precios					−0,075	−0,737
log X_4 Imagen del fabricante					−0,047	−0,463
X_7 Calidad de producto					0,118	1,163

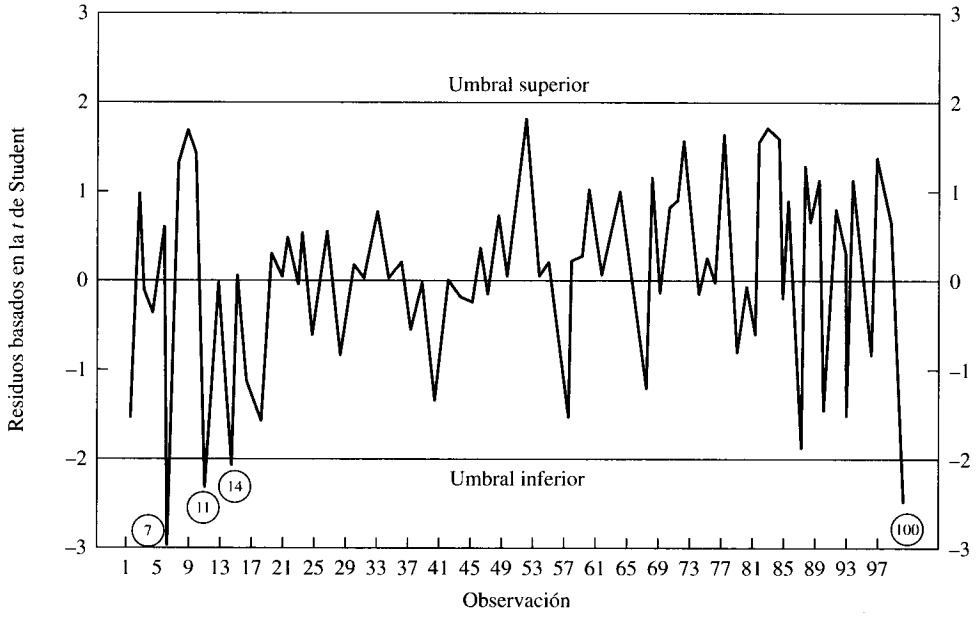


FIGURA 4.13. Gráfico de los residuos basados en la *t* de Student.

El examen de los residuos puede realizarse también a través de gráficos de regresión parcial (véase Figura 4.11). Estos gráficos nos ayudan a identificar situaciones influyentes para cada relación variable independiente-dependiente. Como se observa en la Figura 4.11, puede indentificarse un conjunto de puntos aislados y diferenciados (observaciones 7, 11, 14, 100) para las variables X_3 y X_6. Estos puntos no están bien representados por la relación y por tanto podrían afectar también a la relación parcial. En el Apéndice 4A se discute un análisis más detallado para averiguar si cualquiera de estas observaciones puede clasificarse como una observación influyente.

Quinto paso: Interpretación del resultado teórico

Con la estimación del modelo finalizada, el valor teórico de la regresión especificado y con los contrastes de diagnóstico aplicados que confirman la conveniencia de los resultados, podemos ahora examinar nuestra ecuación de predicción, que incluye X_3, X_5 y X_6. La sección de la Tabla 4.13 titulada «variables de la ecuación» ofrece la ecuación de predicción a partir la columna titulada «coeficiente». Desde esta columna, leemos el término constante (–6,520) y los coeficientes (3,376, 7,621 y 1,406) para las variables 3, 5 y 6, respectivamente. La ecuación predictiva se escribiría:

$$Y = -6{,}520 + 3{,}376X_3 + 7{,}621X_5 + 1{,}406X_6$$

Con esta ecuación, el nivel de fidelidad esperado de cualquier cliente podría calcularse si HATCO conociera sus valoraciones. A efectos ilustrativos, supongamos que un cliente evalúa a HATCO con un valor de 4,0 para cada una de estas tres medidas. El nivel de fidelidad al producto para ese consumidor sería:

$$\begin{aligned}\text{Nivel previsto de fidelidad al producto} &= -6{,}520 + 3{,}376(4{,}0) + 7{,}621(4{,}0) + 1{,}406(4{,}0)\\ &= -6{,}520 + 13{,}504 + 30{,}484 + 5{,}624\\ &= 43{,}902\end{aligned}$$

Además de proporcionar bases para predecir los niveles de fidelidad al producto, los coeficientes de regresión proporcionan también un medio de evaluación de la importancia relativa de las variables individuales en la predicción conjunta del producto. En esta situación, todas las variables están expresadas en la misma escala y pueden realizarse comparaciones directas. Pero en muchos casos los coeficientes beta se usan para comparaciones entre variables independientes. En la Tabla 4.13, los coeficientes beta se enumeran en la columna titulada «coeficiente de regresión estandarizado». El investigador puede hacer comparaciones directas entre las variables para averiguar su importancia relativa en el valor teórico de la regresión. En nuestro ejemplo, X_5 (servicio global) era la más importante, seguida de cerca por X_3 (flexibilidad de precios). La tercera variable independiente, X_6 (imagen de los vendedores) tenía una importancia notablemente baja. Esto explica el escaso aumento de varianza explicada y la reducida correlación univariante con el uso de producto. Aunque significativa, X_6 no merece la atención que debe dirigirse a las otras dos variables independientes.

Medidas del grado y efecto de la multicolinealidad

En cualquier interpretación del valor teórico de la regresión, el investigador debe ser consciente de la influencia de la multicolinealidad. Como se ha discutido previamente, variables altamente colineales pueden distorsionar los resultados sustancialmente o hacerlos muy inestables y, por tanto, no generalizables. Encontramos dos medidas para contrastar la influencia de la colinealidad: (1) calculando la tolerancia y valores VIF, y (2) utilizando los índices de condición y descomponiendo la varianza del coeficiente de regresión. El valor de tolerancia es 1 menos la proporción de la varianza de la variable explicada por las otras variables independientes. Por tanto, un elevado valor de tolerancia indica una reducida colinealidad, y valores de tolerancia cercanos a cero indi-

TABLA 4.15. Contrastes para la multicolinealidad: Valoración de los valores VIF y de tolerancia

Variable	*Tolerancia*	*VIF*
X_3 Flexibilidad de precio	0,99287009	1,00718111
X_5 Servicio global	0,93639766	1,06792236
X_6 Imagen del personal de ventas	0,93946418	1,06443654

can que la variable está casi totalmente tenida en cuenta por las otras variables. El factor de inflación de la varianza (VIF) es el recíproco del valor de tolerancia; así que buscamos valores bajos de VIF como indicación de reducidas correlaciones entre variables. En nuestro ejemplo, los valores de tolerancia superan todos el valor 0,93, indicando niveles de colinealidad muy reducidos (véase Tabla 4.15). De la misma forma, los valores del VIF están muy próximos a 1,0. Estos resultados indican que la interpretación de los coeficientes del valor teórico de la regresión no deberían verse afectados negativamente por la multicolinealidad. Una segunda aproximación para identificar la multicolinealidad y sus efectos es mediante la descomposición del coeficiente de la varianza. Se recomienda al investigador explorar esta técnica y las ideas adicionales que se ofrecen sobre la interpretación de la ecuación de regresión. Los detalles sobre este método se discuten en el apéndice de este capítulo.

Mientras que la multicolinealidad no tiene una influencia sustancial en el valor teórico de la regresión estimada, sí lo tiene en la composición del valor teórico. Después de X_5 (la primera variable incluida en el valor teórico de la regresión), la segunda variable con una correlación más alta con la variable dependiente es X_1. Sin embargo, X_1 tiene también un elevado nivel de colinealidad (0,61) con X_5. Dado que X_5 entró en el valor teórico de la regresión en primer lugar, según el procedimiento por pasos, no existe una varianza única suficiente en X_1 para justificar su inclusión. Por tanto, sólo X_5 entraría en el valor teórico de la regresión. Sin embargo, sería sustancialmente incorrecto inferir de estos resultados que X_1 no tiene efectos sobre la fidelidad al producto cuando de hecho era la variable independiente con la segunda correlación bivariante más elevada con la variable dependiente. La interpretación correcta sería que X_5 o X_1 muestran una elevada influencia, pero que la igualdad de sus efectos sobre el nivel de fidelidad al producto (elevada colinealidad) dicta que sólo se necesita una de ellas para el proceso de regresión. El investigador nunca debe aceptar que un procedimiento de estimación dicte la interpretación de los resultados, sino que debe ser consciente de los resultados de la interpretación que acompañan cada procedimiento de estimación. Por ejemplo, si las siete variables independientes hubieran entrado en el valor teórico de la regresión, el investigador habría tenido que enfrentarse con los efectos de la colinealidad sobre la interpretación de los coeficientes de X_5 y X_1 pero de forma diferente que si se hubiera utilizado el procedimiento por pasos.

Sexto paso: Validación de los resultados

La tarea final a la que se enfrenta el investigador consiste en el proceso de validación del modelo de regresión. El interés primordial de este proceso es asegurar que los resultados son generalizables a la población y que no son específicos de la muestra utilizada para la estimación. La aproximación más directa para la validación es obtener otra muestra de la población y evaluar la correspondencia de los resultados obtenidos de las dos muestras. En ausencia de muestras adicionales, se puede evaluar la validez de los resultados de varias formas. La primera consiste en el examen del valor del R^2 ajustado. En esta situación, el valor del R^2 ajustado es 0,761 (en comparación con un valor del R^2 de 0,768; véase Tabla 4.14), lo que indica que el modelo estimado no está sobreajustado en relación con la muestra y mantiene un adecuado ratio de observaciones respecto a variables en el valor teórico.

Una segunda forma es dividir la muestra en dos submuestras, estimando el modelo de regresión de cada submuestra y comparar los resultados. La Tabla 4.16 contiene los resultados globales por pasos más los resultados a partir de los modelos estimados por pasos de las dos submuestras de 50 observaciones cada una. La comparación del ajuste del modelo conjunto demuestra un elevado nivel de similitud de los resultados en términos del R^2, el R^2 ajustado y el error estándar de la estimación. Pero al comparar el coeficiente individual, aparece una diferencia. En la muestra uno, X_6 no entra en los resultados por pasos mientras que sí lo hace en la muestra dos y en la conjunta. La omisión de X_6 en una de la submuestras confirma que se trataba de un predictor marginal, como indicaba el bajo nivel de beta y los valores t del modelo conjunto.

TABLA 4.16. Validación con división de la muestra de la estimación por etapas

Componente del modelo	*Global* (n = 100)	*Muestra* 1 (n = 50)	*Muestra* 2 (n = 50)
Variables independientes			
X_3 Flexibilidad de precio			
Coeficiente de regresión	3,376	3,108	3,632
Coeficiente beta	0,521	0,506	0,529
Valor t	10,562	6,803	8,439
X_5 Servicio global			
Coeficiente de regresión	7,621	8,278	7,037
Coeficiente beta	0,637	0,710	0,574
Valor t	12,547	9,555	8,954
X_6 Imagen del personal de ventas			
Coeficiente de regresión	1,406	No incluido	2,447
Coeficiente beta	0,121		0,200
Valor t	2,378		3,166
Ajuste del modelo			
R^2	0,768	0,741	0,826
R^2 corregido	0,761	0,730	0,814
Error estándar de la estimación	4,394	4,764	3,816

Evaluación de modelos de regresión alternativos

El modelo de regresión por pasos examinado en la discusión anterior proporcionó una evaluación sólida del problema de investigación tal como se formuló, pero el investigador está siempre dispuesto a evaluar modelos de regresión alternativos en busca de una capacidad explicativa adicional y la confirmación de resultados previos. En esta sección, examinamos dos modelos de regresión adicionales: un modelo que incluye la totalidad de las siete variables independientes en un enfoque de confirmación, y un segundo modelo que añade una variable no métrica (X_8) a través de una variable ficticia.

Modelo de regresión confirmatorio

Una primera alternativa al método de estimación de la regresión por pasos es la aproximación confirmatoria, en la cual el investigador especifica las variables independientes que se van a incluir en la ecuación de regresión. De esta forma, el investigador mantiene un pleno control sobre el valor téorico de la regresión en términos tanto de predicción como de explicación. Esta aproximación es especialmente adecuada en situaciones de réplica de esfuerzos de investigación previos o con propósitos de validación.

En esta situación, la perspectiva confirmatoria implica la inclusión de las siete medidas perceptivas como variables independientes. Éstas son las mismas variables consideradas en el proce-

so de estimación por pasos, pero en este caso todas entran directamente en la ecuación de regresión a un mismo tiempo. Aquí, el investigador puede juzgar los efectos potenciales de la multicolinealidad al seleccionar las variables independientes y el efecto sobre el ajuste global del modelo de incluir las siete variables.

Los resultados de la Tabla 4.17 son similares a los resultados finales a los que se llegaron mediante la estimación por pasos (véase Tabla 4.13), con dos excepciones notables. En primer lugar, a pesar de incluir más variables independientes, el ajuste global del modelo empeora, mientras que el coeficiente de determinación crece (de 0,768 a 0,775) debido a las variables independientes adicionales, el R^2 ajustado disminuye (de 0,761 a 0,758). Esto apunta a la inclusión de varias variables independientes en la ecuación de regresión que no eran significativas. Aunque contribuyen al R^2 global, restan del R^2 ajustado. Esto ilustra el papel del R^2 ajustado al comparar valores teóricos de la regresión con diferente número de variables independientes. Otro indicador del ajuste global más pobre del modelo confirmatorio es el aumento de SEE (de 4,394 a 4,424). Esto ilustra que el R^2 global no debería ser el único criterio de capacidad predictiva debido a que está influido por muchos factores, uno de ellos el número de variables independientes.

La otra diferencia es el valor teórico de la regresión, donde la multicolinealidad afecta al número y relevancia de las variables significativas. En primer lugar, sólo dos variables (X_3 y X_5) son estadísticamente significativas, mientras que el modelo por pasos contiene una tercera variable (X_6). En el modelo por pasos, X_6 era la variable menos significativa, con un valor de t de 2,378. Cuando se utiliza la aproximación confirmatoria, la multicolinealidad de X_6 con X_4, que ahora está in-

TABLA 4.17. Resultado del ejemplo: Resultados de la regresión múltiple utilizando un enfoque confirmatorio con las siete variables independientes

R Múltiple	0,880
R^2 Múltiple	0,775
R^2 Ajustado	0,758
Error estándar de la estimación	4,424

Análisis de la varianza

	Suma de los cuadrados	*gl*	*Cuadrado media*	*Ratio F*
Regresión	6.198,68	7	885,53	45,25
Residuos	1.800,32	92	19,57	

Variables en la ecuación

Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente (beta) de regresión estandarizado*	*Valor t*	*Significación estadística*
Constante	–10,187	4,977		–2,047	0,044
X_1 Velocidad de entrega	–0,058	2,013	–0,008	–0,029	0,977
X_2 Nivel de precios	–0,697	2,090	–0,093	–0,333	0,740
X_3 Flexibilidad de precio	3,368	0,411	0,520	8,191	0,000
X_4 Imagen del fabricante	–0,042	0,667	–0,005	–0,063	0,950
X_5 Servicio global	8,369	3,918	0,699	2,136	0,035
X_6 Imagen del personal de ventas	1,281	0,947	0,110	1,352	0,180
X_7 Calidad de producto	0,567	0,355	0,100	1,595	0,114

cluida en la ecuación de regresión, disminuye la influencia única de X_6 resultando en un coeficiente no significativo. En segundo lugar, la importancia y significación del resto de variables también disminuye en el modelo confirmatorio. Los valores de t y la significación estadística disminuyen tanto para X_3 como para X_5, aunque se mantengan significativas. Esto es resultado de su multicolinealidad con las variables no significativas. Finalmente, la influencia de la multicolinealidad entre X_1 y X_5 impide la inclusión de X_1 en el modelo por pasos a pesar de tener la segunda mayor correlación con la variable dependiente. Por tanto, la multicolinealidad también está reflejada en el modelo confirmatorio por el coeficiente no significativo de X_1 debido a su multicolinealidad con X_3 y X_5 en la ecuación.

La aproximación confirmatoria proporciona al investigador control sobre el valor teórico de la regresión, pero al coste de una ecuación de regresión con menor capacidad predictiva y explicativa si el investigador no examina con atención los resultados. Las aproximaciones confirmatoria y secuencial cuentan con importantes puntos débiles y fuertes que deben considerarse en su uso, pero el investigador prudente empleará ambas para aprovechar las ventajas de cada una.

Inclusión de una variable independiente no métrica

La discusión previa se centró en el método de estimación confirmatorio como alternativa para incrementar posiblemente la capacidad predictiva y explicativa, pero el investigador también debe considerar la posible mejora al añadir variables independientes no métricas. Como se discutió en la sección previa y en el Capítulo 2, las variables no métricas no pueden incluirse directamente en la ecuación de regresión, sino que, en su lugar, deben estar representadas por una serie de variables nuevas creadas, denominadas variables ficticias. En este ejemplo, la variable de tamaño de la empresa (X_8), que cuenta con las dos categorías de grandes y pequeñas empresas, se añadirá a la ecuación de regresión. La variable ya está codificada de forma conveniente para las grandes empresas como un 1 y para las pequeñas como un 0. La variable puede incluirse directamente en la ecuación de regresión y representará la diferencia en la variable dependiente entre grandes y pequeñas empresas. Concretamente, debido a que las grandes empresas tienen el valor 1, las pequeñas empresas actúan como categoría de referencia. El coeficiente de regresión se interpretará como el valor de las grandes empresas comparado con el de las pequeñas. Un coeficiente positivo indicará que las grandes empresas cuentan con un nivel de fidelidad mayor al de las pequeñas empresas, mientras que un coeficiente nagativo indicará que las pequeñas empresas cuentan con un mayor nivel de fidelidad.

La Tabla 4.18 contiene los resultados de añadir X_8 a los resultados finales por pasos. El examen de los estadísticos de ajuste global indica una mejora sustancial, con todas las medidas (el R^2, el R^2 ajustado y SEE) mayores respecto al modelo por pasos (véase Tabla 4.13). Esto viene apoyado por la significación estadística del coeficiente de regresión de X_8 a un nivel que excede el 0,05. El valor positivo del coeficiente (3,852) indica que las grandes empresas, dadas sus características sobre las otras tres variables independientes en la ecuación, tienen un nivel de fidelidad al producto que es un 3,8 por ciento mayor, y que ésta es una diferencia estadísticamente significativa. La adición de X_8 mejora tanto la predicción como la explicación en la pregunta de la investigación. Esto ilustra la forma en que el investigador puede añadir variables no métricas a las variables métricas en el valor teórico de la regresión y mejorar tanto la explicación como la predicción.

Una visión gerencial de los resultados

Los resultados de la regresión, incluyendo la evaluación complementaria del modelo confirmatorio y la adición de la variable no métrica, ayudan a responder a la pregunta básica de la investigación: ¿qué factores afectan a la fidelidad al producto? Al formular una respuesta, el investigador debe considerar dos aspectos: la predicción y la explicación. En términos de predicción, todos los modelos de regresión alcanzan altos niveles de capacidad predictiva. La cantidad de varianza explicada excede el 75 por ciento y la tasa de error esperada de cualquier predicción es del ± 9 por

TABLA 4.18. Resultado del ejemplo: Resultados de la regresión múltiple añadiendo X_8, tamaño de empresa como una variable independiente con la utilización de una variable ficticia

R Múltiple	0,890
R^2 Múltiple	0,793
R^2 Ajustado	0,784
Error estándar de la estimación	4,177

Análisis de la varianza

	Suma de los cuadrados	*gl*	*Cuadrado media*	*Ratio F*
Regresión	6341,826	4	1585,457	90,889
Residuos	1657,174	95	17,444	

Variables en la ecuación

Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente (beta) de regresión estandarizado*	*Valor t*	*Significación estadística*
Constante	–16,335	4,254		–3,840	0,000
X_3 Flexibilidad de precio	4,245	0,399	0,655	10,630	0,000
X_5 Servicio global	8,055	0,592	0,673	13,613	0,000
X_6 Imagen del personal de ventas	1,462	0,562	0,125	2,602	0,011
X_8 Tamaño de la empresa (grande)	3,852	1,149	0,211	3,353	0,001

ciento. En este contexto de investigación, estos niveles, aumentados por los resultados que apoyan la validación del modelo, proporcionan los mayores niveles de seguridad en los modelos de regresión como fundamento para desarrollar estrategias de empresa.

En términos de explicación, todos los modelos estimados llegan esencialmente a los mismos resultados: dos influencias destacadas (la flexibilidad de precios y el servicio global) y una influencia algo menor (imagen de los vendedores) sobre la fidelidad al producto. Incrementos en cualesquiera de estas tres variables resultarán en aumentos correspondientes de la fidelidad al producto. Por ejemplo, un aumento de un punto en la percepción del cliente sobre el servicio global resultará en un incremento medio de al menos un 8 por ciento en la fidelidad al producto. Se comprueban resultados similares para las otras dos variables. Además, al menos una característica de la empresa, el tamaño de la empresa, demuestra un efecto significativo sobre la fidelidad al producto. Las grandes empresas tienen niveles de fidelidad casi un 4 por ciento mayores que las pequeñas empresas. Estos resultados proporcionan a la gestión un marco para desarrollar estrategias que mejoren los niveles de fidelidad al producto. Las acciones dirigidas a incrementar las percepciones de HATCO pueden justificarse a la luz de los correspondientes aumentos de la fidelidad al producto.

Antes de desarrollar cualesquiera conclusiones o planes de empresa a partir de estos resultados, el investigador debería darse cuenta, sin embargo, de que dos de las influencias (la flexibilidad de precios y la imagen de los vendedores) están incluidas en las dos dimensiones perceptivas identificadas en el Capítulo 3. Estas dimensiones, que representan medidas amplias de las percepciones de los clientes sobre HATCO, deberían por tanto ser consideradas en las conclusiones. Establecer que *sólo* estas tres variables específicas influyen sobre la fidelidad al producto sería una falsa afirmación de los patrones más complejos de colinealidad entre las variables. Por tanto, estas variables se ven mejor como representativas de las dimensiones perceptivas, debiendo considerar también las otras variables en cada dimensión en cualesquiera conclusiones derivadas de estos

resultados. Ahora la gestión cuenta con un análisis objetivo que confirma no sólo la influencia concreta de variables clave, sino también las dimensiones perceptivas que deben tenerse en cuenta en cualquier plan de la empresa que incluya estrategias dirigidas a influir en la fidelidad al producto.

Resumen

Este capítulo presenta una introducción simplificada a la lógica y conceptos fundamentales subyacentes en el análisis de regresión múltiple. Pone de manifiesto que los análisis de regresión múltiple pueden describir y predecir la relación entre dos o más variables. Además, el análisis de regresión múltiple, que puede ser utilizado para examinar el poder explicativo total o su incremento de muchas variables, representa una gran mejora respecto del enfoque del análisis secuencial necesario con las técnicas univariantes. Pueden utilizarse tanto las técnicas simultáneas como las de pasos, y bajo ciertas circunstancias pueden incluirse en la ecuación de regresión variables no métricas codificadas como ficticias. Finalmente, hemos visto que existen numerosas técnicas de diagnóstico para contrastar tanto los supuestos subyacentes al análisis de la regresión como para la existencia de casos que ejercen una influencia indebida en las predicción de la ecuación resultante. Este capítulo ha proporcionado una presentación fundamental de cómo opera la regresión y lo que se puede conseguir con ella. Familizarizarse con los conceptos presentados en este capítulo ayudará al lector a entender mejor las presentaciones técnicamente más complejas y detalladas de otros textos además de proporcionar un fundamento para el análisis de regresión que el investigador deba llevar a cabo.

Cuestiones de revisión

1. ¿Cómo explicaría la importancia relativa de las variables independientes utilizadas en una ecuación de regresión?
2. ¿Por qué es importante examinar el supuesto de linealidad cuando se utiliza la regresión?
3. ¿Cómo puede corregirse o tenerse en cuenta la linealidad en la ecuación de regresión?
4. ¿Podría encontrar una ecuación de regresión que fuese aceptable en términos de significación estadística y, sin embargo, no ofrecer un valor interpretable para la gestión práctica?
5. ¿Cuál es la diferencia de interpretación entre los coeficientes de regresión asociados con variables independientes de escala-intervalo y variables independientes codificadas como ficticias (0,1)?
6. ¿Cuáles son las diferencias entre las variables independientes interactivas y correlacionadas? ¿Afecta alguna de estas diferencias a su interpretación de la ecuación de la regresión?
7. ¿Deben omitirse siempre los casos influyentes? Ofrecer ejemplos de ocasiones en las deberían omitirse y en las que no.

Bibliografía

1. Barnett, V., and T. Lewis (1984), Outliers in Statistical Data, 2d ed. New York: Wiley.
2. Belsley, D. A., E. Kuh, and R. E. Welsch (1980), Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.
3. BMDP Statistical Software, Inc. (1991), SOLO Power Analysis. Los Angeles: BMDP.
4. Box, G. E. P., and D. R. Cox (1964), "An Analysis of Transformations." Journal of the Royal Statistical Society B 26: 211–43.
5. Cohen, J., and P. Cohen (1983), Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 2d ed. Hillsdale, N.J.: Lawrence Erlbaum Associates.
6. Daniel, C., and F. S. Wood (1980), Fitting Equations to Data, 2d ed. New York: Wiley-Interscience.

7. Jaccard, J., R. Turrisi, and C. K. Wan (1990), Interaction Effects in Multiple Regression. Beverly Hills, Calif.: Sage Publications.
8. Johnson, R. A., and D. W. Wichern (1982), Applied Multivariate Statistical Analysis. Upper Saddle River, N.J., Prentice Hall.
9. Mason, C. H., and W. D. Perreault, Jr. (1991), "Collinearity, Power, and Interpretation of Multiple Regression Analysis." Journal of Marketing Research 28 (August): 268–80.
10. Mosteller, F., and J. W. Tukey (1977), Data Analysis and Regression. Reading, Mass.: Addison-Wesley.
11. Neter, J., W. Wassermann, and M. H. Kutner (1989), Applied Linear Regression Models. Homewood, Ill.: Irwin.
12. Rousseeuw, P. J., and A. M. Leroy Robust (1987), Regression and Outlier Detection. New York: Wiley.
13. Seer, G. A. F. (1984), Multivariate Observations. New York: Wiley.

APÉNDICE 4A

Validación avanzada en el análisis de regresión múltiple

OBJETIVOS DE APRENDIZAJE

- Comprender cómo el índice de condicionamiento y la matriz de descomposición de la varianza del coeficiente de regresión aislan los efectos, si los hay, de la multicolinealidad de los coeficientes de regresión estimados.
- Identificar aquellas variables con niveles inaceptables de colinealidad o multicolinealidad.
- Identificar las observaciones con un impacto desproporcionado sobre los resultados de la regresión múltiple.
- Aislar las observaciones influyentes y evaluar las relaciones cuando se eliminan las observaciones influyentes.

Probablemente, la técnica estadística más utilizada es la regresión múltiple, y la regresión múltiple ha dirigido el movimiento hacia una mayor utilización de las técnicas multivariantes. Al desplazarse de la regresión simple a la múltiple, la mayor potencia analítica de la forma multivariante requiere también diagnósticos adicionales para tratar con las correlaciones entre las variables y las observaciones con un impacto sustancial sobre los resultados. Este apéndice describe dos técnicas avanzadas de diagnóstico para evaluar: (1) el impacto de la multicolinealidad y (2) la identificación de las observaciones influyentes y su impacto sobre el análisis de regresión múltiple. Mientras que el Capítulo 4 trata con los diagnósticos básicos de estos conceptos, aquí discutiremos procedimientos más avanzados que se han propuesto recientemente para las situaciones multivariantes. Estos procedimientos no son refinamientos de los procedimientos de estimación, sino que dirigen la cuestión a la interpretación de los resultados obtenidos con multicolinealidad y observaciones influyentes.

Evaluación de la multicolinealidad

Como discutimos en el Capítulo 4, la **colinealidad** y **multicolinealidad** pueden tener diversos efectos perversos sobre la regresión múltiple, tanto en la interpretación de los resultados como en la forma de obtenerlos, como ocurre en la regresión por etapas. La utilización de varias variables como predictores hace necesaria la evaluación de la correlación múltiple entre las variables independientes como medio para evaluar la multicolinealidad. Pero esto no es posible examinando sólo la matriz de correlación (que muestra sólo correlaciones simples entre dos variables). Vamos a exponer ahora un método desarrollado específicamente para diagnosticar la extensión de la multicolinealidad presente y las variables que exhiben una mayor multicolinealidad. Todos los principales programas estadísticos tienen análisis opcionales que proporcionan diagnósticos de la colinealidad.

Un proceso en dos partes

El método tiene dos componentes. El primero es el **índice de condicionamiento**, que representa la colinealidad de las combinaciones de variables en el conjunto de datos (en realidad el tamaño relativo de los **autovalores** de la matriz). El segundo es la **matriz de descomposición de la varianza del coeficiente de regresión**, que muestra la proporción de la varianza de cada coeficiente de regresión (y su variable asociada) atribuible a cada índice de condicionamiento (autovalor). Combinamos éstos en un procedimiento en dos partes:

1. Identificar todos los índices de condicionamiento que estén por encima de un valor límite. Este umbral está comprendido habitualmente en un rango entre 15 y 30, con 30 como el valor más utilizado habitualmente.
2. Para todos los índices de condicionamiento que excedan el umbral, identificar variables con proporciones de la varianza por encima del 90 por ciento. Se indica un problema de colinealidad cuando un índice de condicionamiento identificado en el primer paso, por encima del valor de umbral tiene en cuenta una proporción sustancial de la varianza (0,90 o más) de *dos o más* coeficientes.

El ejemplo que se muestra en la Tabla 4A.1 ilustra este procedimiento básico y muestra los índices de condicionamiento y los valores de descomposición de la varianza. Primero, el umbral de 30 del índice de condicionamiento, seleccionamos tres de estos índices (u_4, u_5 y u_6). Segundo, los coeficientes que exceden el umbral del 0,90 para estos tres índices de condicionamiento son b_1 y b_3 con u_6; b_2, b_4 y b_5 con u_4; y b_6 con u_5 (véanse los valores subrayados en la Tabla 4A.1). Sin embargo, u_5 tiene sólo un único valor (b_6) asociado con él; por tanto no se observa colinealidad para este coeficiente. Como resultado, intentaríamos remediar las correlaciones significativas entre los dos conjuntos de variables: (V_1, V_3) y (V_2, V_4, V_5).

TABLA 4A.1. Análisis hipotético de la descomposición de la varianza del coeficiente con índices de condicionamiento

Índice de condición (u_j)	*Proporción de la varianza del coeficiente:* b_1	b_2	b_3	b_4	b_5	b_6
1,0 u_1	0,003	0,001	0,000	0,003	0,000	0,000
4,0 u_2	0,000	0,021	0,005	0,003	0,000	0,000
16,5 u_3	0,000	0,012	0,003	0,010	0,000	0,001
45,0 u_4	0,001	<u>0,963</u>	0,003	<u>0,972</u>	<u>0,983</u>	0,000
87,0 u_5	0,003	0,002	0,000	0,009	0,015	<u>0,988</u>
122,0 u_6	<u>0,991</u>	0,001	<u>0,987</u>	0,003	0,002	0,011

Ejemplo de evaluación de la multicolinealidad

En el Capítulo 4, discutíamos el uso de la regresión múltiple en la predicción del nivel de uso de los productos de HATCO entre sus clientes. El procedimiento por pasos identificó tres predictores significativos: X_3, X_5 y X_6. Sin embargo, antes de aceptar estos resultados de la regresión como válidos, debemos examinar el grado de multicolinealidad y sus efectos sobre los resultados. Para hacerlo, empleamos los índices de condicionamiento y la descomposición de la varianza de los coeficientes y realizaremos comparaciones con el **factor de inflación de la varianza (VIF)** y los valores de **tolerancia**.

Como se expuso en el Capítulo 4 y presente también en la Tabla 4A.2, los valores de tolerancia VIF indican una colinealidad sin consecuencias. Ningún valor del VIF excede 10,0 y se observa en los valores de tolerancia que en ningún caso la colinealidad explica más del 10 por ciento de la varianza de cualquier variable independiente. Esta conclusión se ve corroborada cuando empleamos el procedimiento en dos pasos. En primer lugar, examinamos los índices de condicionamiento. No iríamos más allá del primer paso, en la medida en que ningún índice de condicionamiento es mayor que 30,0. Incluso si fuéramos a utilizar un valor umbral de 15 para el índice de condicionamiento, seleccionaríamos sólo u_4 y sólo un coeficiente (la constante) lo sobrepasaría. Por tanto, no podemos encontrar evidencia que apoye la existencia de multicolinealidad en estos resultados de regresión, tal y como se indica por las medidas de tolerancia y VIF.

TABLA 4A.2. Contraste de la multicolinealidad en la regresión múltiple

PARTE A. VALORACIÓN DE LOS VALORES VIF Y TOLERANCIA

Variable	*Tolerancia*	*Factor de la inflación de la varianza (VIF)*
X_3	0,993	1,007
X_5	0,936	1,068
X_6	0,939	1,064

PARTE B. UTILIZACIÓN DE LOS ÍNDICES DE CONDICIONAMIENTO Y DESCOMPOSICIÓN DE LA MATRIZ DE LA VARIANZA DEL COEFICIENTE

			Proporción de la varianza del coeficiente			
Número	*Autovalor*	*Índice de condicionamiento*	*Constante*	X_3	X_5	X_6
1	3,882	1,000	0,001	0,002	0,004	0,005
2	0,060	8,046	0,014	0,110	0,021	0,850
3	0,045	9,246	0,020	0,136	0,909	0,042
4	0,012	17,719	0,965	0,753	0,066	0,103

* Los valores subrayados exceden el nivel de 0,90.

Identificación de observaciones influyentes

En el Capítulo 4, examinamos sólo una aproximación para identificar **observaciones influyentes**, el uso de residuos estandarizados para encontrar los casos atípicos. Como se apuntó entonces, sin embargo, hay casos que según este método pueden clasificarse como influyentes aunque no sean reconocidos como casos atípicos. De hecho, muchas veces una observación de influencia no será identificada como un caso atípico porque ha tenido influencia en la estimación de regresión hasta tal punto que su residuo resulta insignificante. Por tanto, necesitamos examinar procedimientos más específicos para medir la influencia de una observación en los diversos aspectos de la regresión múltiple [2]. En la siguiente exposición, discutiremos un proceso de cuatro pasos para la identificación de atípicos, puntos de apalancamiento y observaciones influyentes. Como se observó previamen-

te, una observación puede pertenecer a una o varias de estas categorías, y en este caso es el juicio del investigador, a la luz de la mejor evidencia disponible, el que elegirá el curso de acción a tomar.

Primer paso: Examen de los residuos

Los residuos son fundamentales en la detección de los incumplimientos de los supuestos del modelo y también juegan un papel importante en la identificación de observaciones que son **casos atípicos** de la variable dependiente. Emplearemos dos métodos de detección: el análisis de residuos y gráficos de regresión parcial.

Análisis de residuos

El **residuo** es la manera principal de clasificar una observación como un caso atípico. El residuo para la *i*-ésima observación se calcula como los valores reales menos los valores de predicción de la variable dependiente, o:

$$\text{Residuo}_i = Y_i - \hat{Y}_i$$

Sin embargo, el residuo y sus múltiples formas se basan realmente en dos procedimientos: los casos utilizados para calcular el valor predictor y la utilización (o no utilización) de algún modo de estandarización. Ya hemos visto cómo calculamos el residuo, pero una segunda forma, el **residuo eliminado**, es diferente del residuo normal porque se omite la *i*-ésima observación cuando se estima la ecuación de regresión utilizada para calcular el valor predictor para aquella observación. Por tanto, cada observación no tiene impacto sobre su propio valor predictor en el residuo eliminado. Se utiliza el residuo eliminado menos frecuentemente, aunque tiene la ventaja de reducir la influencia de la observación sobre su cálculo.

El segundo procedimiento en la definición de un residuo implica la cuestión de estandarizar o no los residuos. Los residuos que no son estandarizados están en la misma escala que la variable dependiente, lo que es útil en cuanto a la interpretación pero no nos proporciona una comprensión de lo que significa demasiado grande o suficiente pequeño para no tenerlo en cuenta. Los **residuos estandarizados** son el resultado de un proceso de creación de una escala común con la división de cada residuo por la desviación estándar de los residuos. Después de la estandarización, los residuos tienen una media de 0 y una desviación estándar de 1. Con un tamaño muestral bastante grande (50 o más), los residuos estandarizados siguen aproximadamente la distribución *t*, de tal forma que los residuos que exceden un umbral como el de 1,96 (el valor crítico *t* al nivel de confianza de 0,05) pueden ser considerados estadísticamente significativos. Las observaciones que no logran el umbral son estadísticamente significativas en su diferencia de 0 y pueden ser considerados como casos atípicos. Esto significa que el valor predictor también es diferente de forma significativa del valor real al nivel de 0,05. Se ha propuesto un test más estricto, que tiene en cuenta la comparación múltiple hecha entre los diversos tamaños muestrales [4].

Una forma especial del residuo estandarizdo es el **residuo basado en la *t* de Student**. En su concepto, es similar al residuo eliminado, pero en este caso, la observación *i*-ésima está eliminada cuando se derive la desviación estándar utilizada para estandarizar el residuo *i*-ésimo. Lo racional es que si una observación es extremadamente influyente, puede ser que no este identificada por los residuos estandarizados normales debido a su impacto sobre el modelo de regresión estimado. El residuo basado en la *t* de Student elimina el impacto del caso sobre el proceso de estandarización y ofrece una medida residual «menos influyente». Esto puede ser evaluado por los mismos criterios que el residuo estandarizado.

Los cinco tipos de residuos que se calculan típicamente con la combinación de las opciones para el cálculo y la estandarización son (1) el residuo normal, (2) el residuo eliminado, (3) el residuo estandarizado, (4) el residuo basado en la *t* de Student y (5) el residuo basado en la *t* de Student

eliminado. Cada tipo de residuo ofrece unas perspectivas únicas tanto de la exactitud de predicción de la ecuación de regresión mediante su designación de casos atípicos como las influencias posibles de la observación acerca de los resultados globales.

Gráficos de regresión parcial

Para representar gráficamente el impacto de los casos individuales, el gráfico de regresión parcial es más efectivo. Dado que la pendiente de la línea de regresión del gráfico de regresión parcial es igual al coeficiente de la variable en la ecuación de regresión, puede verse rápidamente el impacto de un caso aislado sobre la pendiente de la regresión (y el correspondiente coeficiente de la ecuación de regresión). Los efectos de estos casos atípicos sobre los coeficientes de regresión individual quedan representados gráficamente. De nuevo, la mayoría de los programas informáticos tienen la opción del gráfico de regresión parcial, por lo que el investigador sólo necesita observar los casos atípicos aislados del resto de las observaciones. Una comparación de los gráficos de regresión parcial con y sin la observación (observaciones) con consideración influyente puede demostrar su impacto.

Segundo paso: Identificación de punto de apalancamiento mediante los predictores

Nuestro siguiente paso es encontrar aquellas observaciones aisladas del resto de las observaciones sobre una o más variables independientes. A estos casos se les denominan **puntos de apalancamiento** en los que puede «apalancar» la relación en su dirección debido a sus diferencias respecto del resto de las observaciones (véase Capítulo 4 para una descripción general de los puntos de apalancamiento).

Matriz sombrero

Cuando sólo tenemos dos variables predictor, una forma de observar aquellas observaciones diferentes del resto es el gráfico bidimensional, en cuyos ejes situaremos las dos variables. Aunque, cuando se incluyen un mayor número de variables predictor en la ecuación de la regresión, es claramente imposible llevar a cabo esta técnica mediante los métodos univariantes. Sin embargo, podemos utilizar una matriz especial, la **matriz sombrero**, que contiene valores **(valores de sombrero)** para cada observación, que indican el apalancamiento. Los valores sombrero representan los efectos combinados de todas las variables independientes para cada caso.

Los valores sombrero (encontrados en la diagonal de la matriz sombrero) miden dos aspectos de la influencia. En primer lugar, para cada observación, el valor sombrero es una medida de la distancia de la observación desde el centro medio del resto de las observaciones de las variables independientes. (Parecido a la distancia de Mahalanobis, véase más adelante.) En segundo lugar, los valores de la diagonal más larga indican también qué observación da una ponderación desproporcionada al determinar el valor de la predicción de la variable criterio, minimizando así su residuo. Se trata de una indicación de influencia, dado que la línea de regresión debe estar cercana a esta observación (es decir, fuertemente influenciada) para obtener un residuo pequeño. Esto no es necesariamente «malo», como se explica en el Capítulo 4, cuando las observaciones influyentes están dentro de la pauta general de las observaciones restantes.

¿Qué entendemos por un valor sombrero alto? La gama de valores posibles están entre 0 y 1, y el valor medio es p/n, donde p es el número de predictores (el número de coeficientes más uno por la constante) y n es el tamaño muestral. La norma para situaciones donde p es mayor que 10 y el tamaño muestral excede las 50 observaciones, es elegir aquéllas con un valor de apalancamiento superior a dos veces la media ($2p/n$). Cuando el número de predictores o el tamaño muestral es me-

nor de 10 o el tamaño muestral es menor que 50, se sugiere utilizar tres veces la media. Los programas informáticos más difundidos tienen todos opciones para calcular e imprimir los valores de apalancamiento de cada observación. El investigador debe entonces seleccionar el valor límite apropiado ($2p/n$ o $3p/n$) e identificar observaciones con valores mayores que el límite.

Distancia de Mahalanobis

Una medida comparable es la **distancia de Mahalanobis (D^2)**, que considera sólo la distancia de una observación desde los valores medios de las variables independientes y no el impacto sobre el valor de la predicción. La distancia de Mahalanobis se ha explicado como medio para identificar los casos atípicos (abordado en el Capítulo 2). En este caso es de valor limitado, dado que los valores límites dependen de un número de factores y no es posible encontrar una regla para el cálculo de los valores. Es posible determinar la significación estadística a partir de tablas confeccionadas [1]. Aun incluso sin las tablas, el investigador puede observar los valores e identificar cualquier observación con valores sustancialmente altos respecto al resto de las observaciones. Por ejemplo, si un reducido conjunto de observaciones con los valores de Mahalanobis más elevados son dos o tres veces el siguiente valor más alto, constituiría un corte sustancial en la distribución y sería otro indicador de un posible apalancamiento.

Tercer paso: Diagnóstico de un único caso de observaciones influyentes

Hasta ahora hemos encontrado puntos atípicos de las variables criterio y predictor pero no hemos estimado formalmente la influencia de una única observación de los resultados. En este tercer paso, todos los métodos descansan sobre una proposición común: la medida más directa de influencia implica la eliminación de una o más observaciones y la observación de los cambios en los resultados de regresión en términos de residuos, coeficientes individuales, matriz de coeficientes o ajuste conjunto del modelo. El investigador sólo tiene que examinar los valores y seleccionar aquellas observaciones que exceden el valor especificado. Ya hemos abordado una de estas medidas, el residuo basado en la t de Student eliminado, pero ahora exploraremos otras medidas apropiadas para la diagnosis de los casos individuales.

Influencias sobre coeficientes individuales

El impacto de la eliminación de una única observación sobre cada coeficiente de la regresión se observa mediante el **DFBETA** y su versión estandarizada, el **SDFBETA**. Calculado como el cambio en el coeficiente cuando la observación es eliminada, el DFBETA es el efecto relativo de una observación sobre cada coeficiente. Las pautas generales para identificar valores particularmente altos de SDFBETA sugieren que se aplique el umbral de $\pm 1{,}0$ o $\pm 2{,}0$ a los tamaños muestrales pequeños, mientras $\pm 2\sqrt{n}$) debería ser utilizado para medianos y grandes conjuntos de datos.

Medidas globales de influencia

La distancia de Cook (D_i), se considera la medida más representativa de la influencia global. Captura el impacto de una observación a través de dos fuentes: el tamaño de los cambios en los valores previstos cuando se omite el caso (residuos atípicos basados en la t de Student), así como la distancia de la observación desde el resto de las observaciones (apalancamiento). La norma es identificar observaciones con una distancia de Cook de 1,0 o mayor, aunque el umbral de $4/(n-k-1)$, donde k es el número de variables independientes, se considera como una medida más conserva-

dora en muestras pequeñas para el uso con series de datos más grandes. Incluso si ninguna observación excede este límite, hay que prestar una atención especial si un conjunto pequeño de observaciones tiene valores sustancialmente más elevados que el resto de las observaciones.

Una medida similar es el **COVRATIO**, que estima los efectos de la observación sobre la eficiencia del proceso de estimación. Específicamente, COVRATIO representa el grado al que una observación tiene impacto sobre los errores estándar de los coeficientes de regresión. Difiere del DFBETA y SDFBETA en que considera conjuntamente a los coeficientes en lugar de hacerlo individualmente.Un umbral puede ser establecido a $1 \pm 3p/n$. Los valores por encima del umbral de $1 + 3p/n$ hace que el proceso de estimación sea más eficiente, mientras que aquellos con menos de $1 - 3p/n$ resta valor a la eficiencia de estimación. Esto permite al COVRATIO actuar como otro indicador de observaciones que tienen una influencia sustancial de forma positiva y negativa sobre el conjunto de coeficientes.

Una tercera medida es **SDFFIT**, el grado en el cual los valores ajustados cambian cuando se elimina el caso. Un valor de corte de $2\sqrt{(k+1)/(n-k-1)}$ ha sido sugerido para detectar una influencia sustancial. Aun cuando la distancia de Cook y SDFFIT son medidas de ajuste global, deben ser complementadas por las medidas del primer y segundo paso que nos permita determinar si la influencia procede de los residuos, el apalancamiento o de ambos. También se dispone de una versión sin estandarizar (**DFFIT**).

Cuarto paso: Selección y tratamiento de observaciones influyentes

La identificación de observaciones influyentes es más un proceso de convergencia por múltiples métodos que una mera dependencia de una única medida. Dado que ninguna medida aislada representa todas las dimensiones de la influencia, se trata de un asunto de interpretación, aunque estas medidas identifican normalmente un conjunto pequeño de observaciones. Al seleccionar las observaciones con altos valores en las medidas de diagnóstico, habría que identificar en primer lugar todas las observaciones que excedan los valores límites. A continuación, se debería examinar el rango de valores del conjunto de datos que está siendo analizado buscando diferencias importantes entre los valores altos y los restantes datos. Se detectarán algunas observaciones adicionales que deberían ser clasificadas como influyentes.

Después de la identificación, se abren varios cursos de acción. El primero, si el número de observaciones es pequeño y hay un argumento justificable para su exclusión, los casos deberían ser eliminados y reeditarse la ecuación. Si, por el contrario, la eliminación no está justificada, existen diversas técnicas de estimación más «robustas», entre ellas la regresión robusta [3]. Cualquiera que sea la acción que se tome, debería cumplir nuestro objetivo original de hacer el conjunto de datos más representativo de la población efectiva, para asegurar la validez y la generalización.

Ejemplo de la base de datos HATCO

Para ilustrar los procedimientos de diagnóstico de observaciones influyentes, examinaremos el ejemplo de regresión del Capítulo 4. Al identificar cualquier observación como «influyente» (que tiene un efecto desproporcionado sobre los resultados de la regresión), determinaremos también si debería excluirse del análisis. El proceso de cuatro pasos descrito anteriormente es el que se usa en los siguientes pasos.

Primer paso: *Análisis de los residuos*

Nuestro primer análisis consiste en el examen de los residuos y la identificación de cualquier caso atípico (es decir, observaciones con altos residuos que no se predicen bien por la ecuación de re-

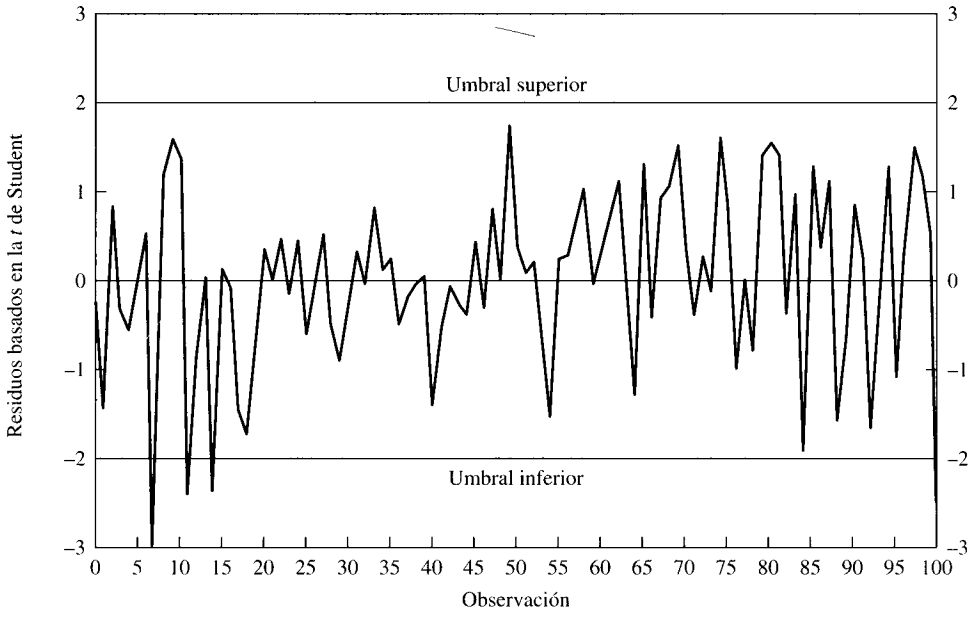


FIGURA 4A.1. Residuos basados en la *t* de Student.

gresión). La Figura 4A.1 muestra los cinco tipos de residuos para las 100 observaciones. Tal y como se puede observar, existe una relación estrecha entre los valores relacionados (es decir, de residuos versus residuos eliminados o residuo estandarizado versus basado en la *t* de Student). Sin embargo, una característica destacada es que conforme aumentan los valores del residuo, crece también la diferencia entre los valores relacionados. Esto indica que el proceso de eliminación de una observación de afectar su propio residuo hace una diferencia. Los residuos basados en la *t* de Student constituyen la forma de residuos más común para denotar los casos atípicos. La Figura 4A.1 demuestra los residuos basados en la *t* de Student para cada observación. Dado que los residuos corresponden a los valores *t*, habiéndose fijado los límites superior e inferior en un intervalo de confianza del 95 por ciento (valor $t = 1{,}96$). Los residuos estadísticamente significativos son aquellos que no entran dentro de esos límites. Cuatro observaciones (7, 11, 14 y 100) tienen residuos estadísticamente significativos y pueden clasificarse como atípicos. Un caso atípico no es necesariamente un punto influyente, ni todos los puntos influyentes tienen que ser casos atípicos. Pero los casos atípicos son importantes en la medida en que representan observaciones no representadas por la ecuación de regresión dado que por una o varias razones, uno de ellos puede tener efectos influyentes en la ecuación que puedan requerir una solución.

El examen de los residuos puede hacerse también a través de los gráficos de regresión parcial (véase Figura 4A.2). Estos gráficos ayudan a identificar observaciones influyentes para cada predictor, relación de la variable criterio. Como se señaló para la Figura 4A.2, puede identificarse un conjunto de puntos separados y diferenciados (observaciones 7, 11, 14, 100) como observaciones influyentes posibles para las tres variables independientes X_3, X_5 y X_6. Estos puntos no están bien representados por la relación y por tanto también podrían afectar hasta cierto punto las coeficientes de regresión estimadas.

Segundo paso: *Identificación de puntos de apalancamiento*

Aunque el análisis de los residuos podía identificar los atípicos, era incapaz de observar el impacto de cada observación sobre las predicciones. Para hacerlo, utilizamos la medida de apalancamiento (véase Tabla 4A.4 y Figura 4A.3). Utilizamos los límites de $2p/n$, dado que el tamaño muestral excede de 50. El valor calculado es $2 \times 4/100 = 0{,}08$. Utilizando este límite, identificamos siete

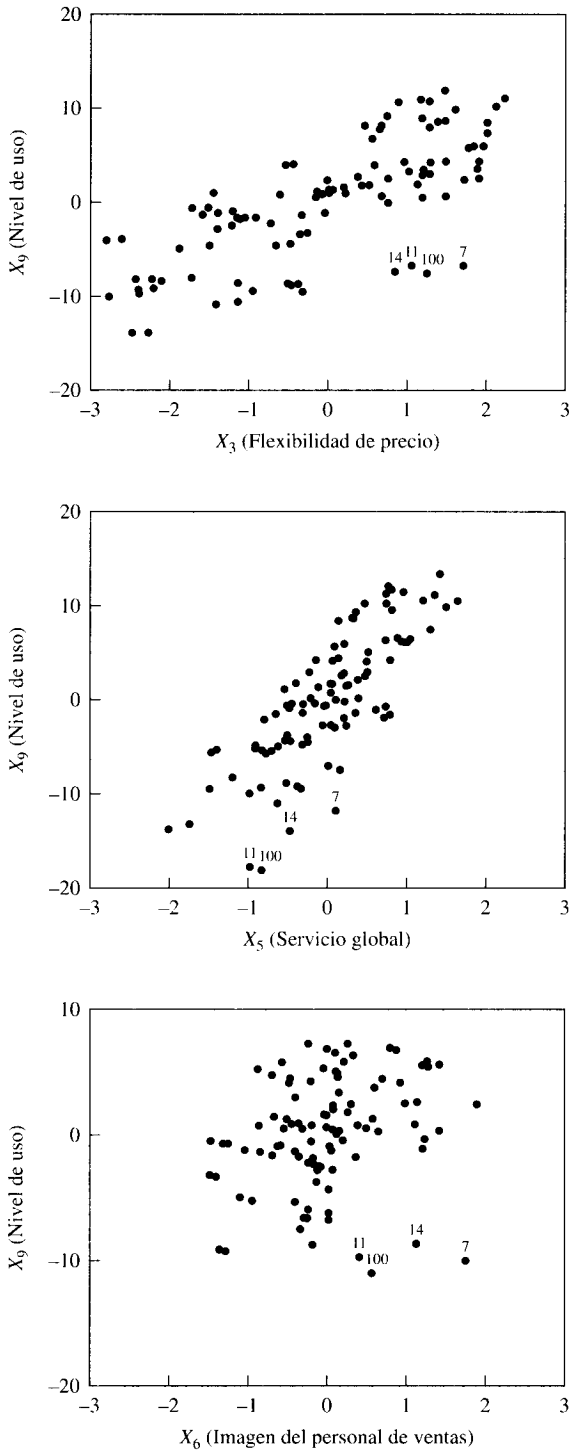


FIGURA 4A.2. Gráficos de regresión parcial estandarizados y observaciones influyentes.

casos (5, 42, 82 y 96) como puntos de apalancamiento. Estos casos no tienen un efecto influyente sobre los resultados, pero como se ha discutido previamente, no todos los puntos influyentes son malos (es decir, no todos distorsionan o se eliminan de la relación). Debemos examinar las observaciones aún más para evaluar qué partes de los resultados quedan afectadas.

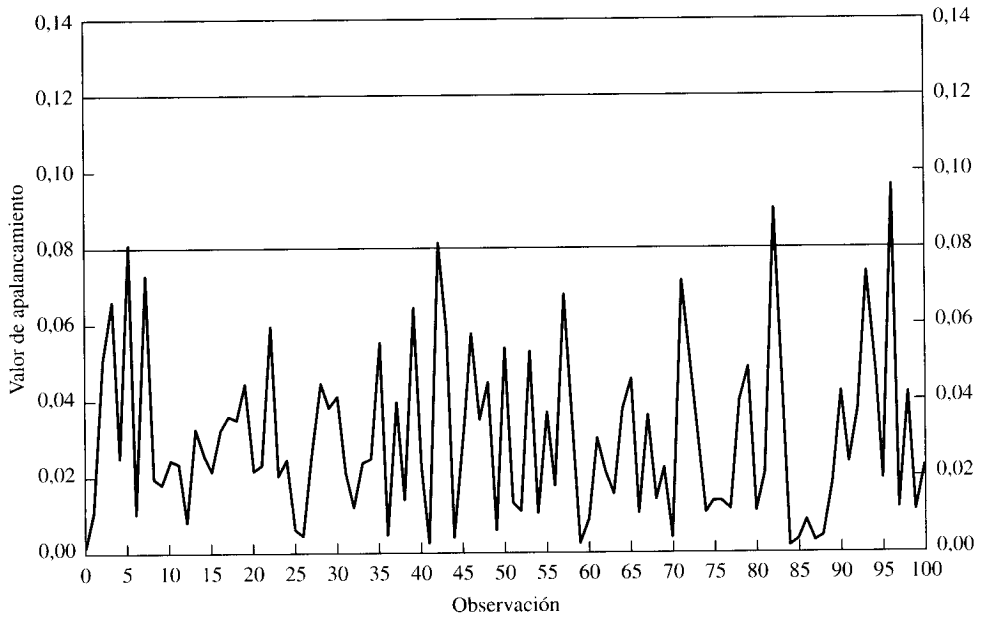


FIGURA 4A.3. Valores de apalancamiento.

Tercer paso: *Análisis de caso único*

Una vez que hemos identificado los casos potencialmente influyentes, complementamos ahora el análisis con medidas que reflejan el impacto sobre partes específicas de los resultados de la regresión. En cada caso, eliminaremos un caso de la estimación de la regresión y observaremos los cambios en los resultados. Nuestra primera medida es el residuo estandarizado eliminado (véase Tabla 4A.3). Como con el análisis de los residuos en el primer paso, de nuevo tenemos cuatro casos significativos pero que se unen en un caso adicional (caso 84). En cada caso, se trata de casos atípicos del modelo de regresión estimado sin su influencia tanto en el cálculo de los valores de predicción como el proceso de estandarización.

El paso siguiente consiste en identificar el impacto sobre los coeficientes estimados de la regresión utilizando los valores SDFBETA (véase Tabla 4A.4 y Figura 4A.4). Dado su formato estandarizado y el tamaño muestral, el umbral será establecido a $\pm 2/\sqrt{n}$ ó 0,08. Mientras que se identifican una cantidad específica de observaciones para cada coeficiente, se identifican varias observaciones que afectan a los coeficientes múltiples (7, 40, 94, 98 y 100). Además, la observación 7 tiene el impacto mayor sobre el constante y los coeficientes para X_3 y X_5, y la observación 11 tiene el mayor impacto sobre el coeficiente para X_5. Podemos combinar estos resultados con otras medidas para identificar la serie final de medidas influyentes.

Disponemos de tres medidas para considerar todos los coeficientes de forma simultánea. La primera medida es la distancia de Cook, que coge tanto los efectos de apalancamiento como el cambio en los residuos (véase Tabla 4A.4 y Figura 4A.5). El umbral de 0,042 calculado [4/(100 – 3 – 1)] identifica de nuevo las cuatro observaciones de 7, 11, 14 y 100. El COVRATIO (véase Tabla 4A.4 y Figura 4A.6), identifica ocho observaciones que contribuyen positivamente al proceso de estimación (3, 5, 39, 42, 71, 82, 93 y 96) mientras que sólo cuatro observaciones restan valor al proceso (7, 11, 14 y 100). Finalmente, la medida SDFFIT (véase Tabla 4A.4 y Figura 4A.7) con un umbral de 0,408 también selecciona cuatro observacioens (7, 11, 14 y 100) como posibles observaciones influyentes.

La Tabla 4A.5 resume los resultados de las medidas diagnósticas para la identificación de las observaciones influenciales. En todas las medidas, cuatro observaciones (7,11,14 y 100) se destacan como puntos potencialmente de negativa influencia debido a su sustancial influencia y a las diferencias respecto del resto de las observaciones. La observación 7 es sustancialmente mayor que

TABLA 4A.3. Comparación de cinco tipos básicos de residuos

Observación	*Valor real*	*Valor de predicción*	*Residuo*	*Residuo eliminado*	*Residuo estandarizado*	*Residuo basado en la t de Student*	*Residuo basado en la t de Student eliminado*
1	32,0	38,299	−6,299	−6,431	−1,434	−1,449	−1,457
2	43,0	39,425	3,575	3,807	0,814	0,840	0,838
3	48,0	49,290	−1,290	−1,397	−0,294	−0,306	−0,304
4	32,0	34,401	−2,401	−2,487	−0,546	−0,556	−0,554
5	58,0	58,268	−0,268	−0,295	−0,061	−0,064	−0,064
6	45,0	42,637	2,363	2,412	0,538	0,543	0,541
7	46,0	58,552	−12,552	−13,687	−2,857	−2,983	−3,115
8	44,0	38,843	5,157	5,313	1,174	1,191	1,194
9	63,0	56,106	6,894	7,092	1,569	1,591	1,604
10	54,0	48,121	5,879	6,088	1,338	1,362	1,368
11	32,0	42,367	−10,367	−10,724	−2,360	−2,400	−2,462
12	47,0	50,580	−3,580	−3,646	−0,815	−0,822	−0,821
13	39,0	38,798	0,202	0,211	0,046	0,047	0,047
14	38,0	48,292	−10,292	−10,675	−2,342	−2,386	−2,447
15	54,0	53,421	0,579	0,598	0,132	0,134	0,133
16	49,0	49,194	−0,194	−0,203	−0,044	−0,045	−0,045
17	38,0	44,236	−6,236	−6,536	−1,419	−1,453	−1,462
18	40,0	47,496	−7,496	−7,847	−1,706	−1,746	−1,765
19	54,0	56,860	−2,860	−3,025	−0,651	−0,669	−0,667
20	55,0	53,421	1,579	1,630	0,359	0,365	0,363
21	41,0	41,049	−0,049	−0,050	−0,011	−0,011	−0,011
22	35,0	33,036	1,964	2,111	0,447	0,463	0,462
23	55,0	55,655	−0,655	−0,676	−0,149	−0,151	−0,151
24	36,0	34,094	1,906	1,974	0,434	0,442	0,440
25	49,0	51,656	−2,656	−2,700	−0,605	−0,609	−0,607
26	49,0	49,999	−0,999	−1,013	−0,227	−0,229	−0,228
27	36,0	33,756	2,244	2,327	0,511	0,520	0,518
28	54,0	56,097	−2,097	−2,219	−0,477	−0,491	−0,489
29	49,0	52,918	−3,918	−4,115	−0,892	−0,914	−0,913
30	46,0	47,339	−1,339	−1,410	−0,305	−0,313	−0,311
31	43,0	41,579	1,421	1,466	0,323	0,329	0,327
32	53,0	53,175	−0,175	−0,179	−0,040	−0,040	−0,040
33	60,0	56,449	3,551	3,673	0,808	0,822	0,820
34	47,0	46,540	0,460	0,476	0,105	0,106	0,106
35	35,0	33,978	1,022	1,094	0,233	0,241	0,240
36	39,0	41,227	−2,227	−2,259	−0,507	−0,510	−0,508
37	44,0	44,774	−0,774	−0,814	−0,176	−0,181	−0,180
38	46,0	46,147	−0,147	−0,151	−0,034	−0,034	−0,034
39	29,0	28,812	0,188	0,203	0,043	0,044	0,044
40	28,0	34,184	−6,184	−6,400	−1,407	−1,432	−1,440
41	40,0	42,467	−2,467	−2,498	−0,562	−0,565	−0,563
42	58,0	58,268	−0,268	−0,295	−0,061	−0,064	−0,064
43	53,0	54,141	−1,141	−1,225	−0,260	−0,269	−0,268
44	48,0	49,654	−1,654	−1,677	−0,376	−0,379	−0,377
45	38,0	36,105	1,895	1,970	0,431	0,440	0,438
46	54,0	55,381	−1,381	−1,482	−0,314	−0,326	−0,324
47	55,0	51,553	3,447	3,606	0,784	0,802	0,801
48	43,0	43,055	−0,055	−0,059	−0,013	−0,013	−0,013
49	57,0	49,426	7,574	7,694	1,724	1,737	1,756
50	53,0	51,417	1,583	1,691	0,360	0,372	0,371
51	41,0	40,682	0,318	0,325	0,072	0,073	0,073

(Continúa)

TABLA 4A.3. *(Continuación)*

Observación	*Valor real*	*Valor de predicción*	*Residuo*	*Residuo eliminado*	*Residuo estandarizado*	*Residuo basado en la t de Student*	*Residuo basado en la t de Student eliminado*
52	53,0	52,114	0,886	0,904	0,202	0,204	0,203
53	50,0	52,300	−2,300	−2,454	−0,523	−0,541	−0,539
54	32,0	38,807	−6,807	−6,946	−1,549	−1,565	−1,577
55	39,0	37,997	1,003	1,052	0,228	0,234	0,233
56	47,0	45,875	1,125	1,156	0,256	0,260	0,258
57	62,0	59,143	2,857	3,098	0,650	0,677	0,675
58	65,0	60,592	4,408	4,612	1,003	1,026	1,027
59	46,0	46,278	−0,278	−0,281	−0,063	−0,064	−0,063
60	50,0	48,731	1,269	1,293	0,289	0,292	0,290
61	54,0	51,075	2,925	3,046	0,666	0,679	0,677
62	60,0	55,209	4,791	4,943	1,090	1,108	1,109
63	47,0	47,725	−0,725	−0,744	−0,165	−0,167	−0,166
64	36,0	41,615	−5,615	−5,893	−1,278	−1,309	−1,314
65	40,0	34,479	5,521	5,845	1,256	1,293	1,297
66	45,0	46,912	−1,912	−1,951	−0,435	−0,440	−0,438
67	59,0	55,172	3,828	4,013	0,871	0,892	0,891
68	46,0	41,465	4,535	4,642	1,032	1,044	1,045
69	58,0	51,521	6,479	6,693	1,474	1,499	1,509
70	49,0	47,598	1,402	1,421	0,319	0,321	0,320
71	50,0	51,771	−1,771	−1,927	−0,403	−0,420	−0,419
72	55,0	53,898	1,102	1,176	0,251	0,259	0,258
73	51,0	51,675	−0,675	−0,703	−0,154	−0,157	−0,156
74	60,0	53,147	6,853	6,991	1,560	1,575	1,588
75	41,0	37,848	3,152	3,226	0,717	0,726	0,724
76	49,0	53,398	−4,398	−4,501	−1,001	−1,013	−1,013
77	42,0	41,922	0,078	0,079	0,018	0,018	0,018
78	47,0	50,438	−3,438	−3,617	−0,782	−0,803	−0,801
79	39,0	33,239	5,761	6,119	1,311	1,351	1,357
80	56,0	49,429	6,571	6,706	1,495	1,511	1,521
81	59,0	53,240	5,760	5,940	1,311	1,331	1,337
82	47,0	48,718	−1,718	−1,909	−0,391	−0,412	−0,410
83	41,0	36,944	4,056	4,316	0,923	0,952	0,952
84	37,0	45,516	−8,516	−8,613	−1,938	−1,949	−1,979
85	53,0	47,452	5,548	5,621	1,263	1,271	1,275
86	43,0	41,544	1,456	1,483	0,331	0,334	0,333
87	51,0	46,211	4,789	4,850	1,090	1,097	1,098
88	36,0	43,035	−7,035	−7,133	−1,601	−1,612	−1,626
89	34,0	36,882	−2,882	−2,959	−0,656	−0,665	−0,663
90	60,0	56,412	3,588	3,786	0,817	0,839	0,838
91	49,0	48,018	0,982	1,015	0,223	0,227	0,226
92	39,0	46,256	−7,256	−7,615	−1,651	−1,692	−1,709
93	43,0	43,757	−0,757	−0,826	−0,172	−0,180	−0,179
94	36,0	30,641	5,359	5,706	1,220	1,259	1,262
95	31,0	35,818	−4,818	−4,961	−1,097	−1,113	−1,114
96	25,0	23,373	1,627	1,820	0,370	0,392	0,390
97	60,0	53,625	6,375	6,512	1,451	1,466	1,475
98	38,0	33,122	4,878	5,143	1,110	1,140	1,142
99	42,0	39,991	2,009	2,051	0,457	0,462	0,460
100	33,0	44,848	−11,848	−12,241	−2,696	−2,741	−2,840
Mínimo	25,0	23,373	−12,552	−13,687	−2,857	−2,983	−3,115
Máximo	65,0	60,592	7,754	7,694	1,724	1,737	1,756

Nota: Los residuos estadísticamente significativos están subrayados.

TABLA 4A.4. Medidas de diagnóstico para identificar observaciones influyentes

Observación	*Distancia de Mahalanobis*	*Distancia de Cook*	*Apalancamiento (valores sombrero)*	*COVRATIO*	*SDFFIT*	*SDFBETA*			
						Constante	X_3	X_5	X_6
1	1,049	0,011	0,011	0,975	–0,211	–0,168	0,102	0,079	0,052
2	5,033	0,011	0,051	1,078	0,213	0,045	–0,088	–0,080	0,162
3	6,564	0,002	0,066	1,125	–0,087	–0,020	0,055	–0,064	0,016
4	2,453	0,003	0,025	1,066	–0,105	–0,075	0,027	0,078	0,008
5	8,040	0,000	0,081	1,147	–0,020	0,014	–0,009	0,000	–0,017
6	1,020	0,002	0,010	1,051	0,078	0,038	–0,001	–0,010	–0,050
7	7,219	0,201	0,073	0,771	–0,937	0,657	–0,402	–0,038	–0,759
8	1,917	0,011	0,019	1,012	0,208	0,167	–0,152	0,011	–0,079
9	1,775	0,018	0,018	0,964	0,272	–0,191	0,172	0,100	0,051
10	2,410	0,016	0,024	0,999	0,258	0,031	–0,148	0,137	0,056
11	2,301	0,050	0,023	0,842	–0,457	–0,007	–0,192	0,341	–0,131
12	0,801	0,003	0,008	1,032	–0,111	0,040	–0,071	–0,008	0,017
13	3,222	0,000	0,033	1,089	0,010	0,005	0,001	–0,003	–0,007
14	2,566	0,053	0,026	0,847	–0,472	0,172	–0,152	0,169	–0,372
15	2,095	0,000	0,021	1,076	0,024	–0,013	0,020	0,000	–0,001
16	3,163	0,000	0,032	1,088	–0,009	0,001	–0,006	0,000	0,005
17	3,541	0,025	0,036	1,000	–0,320	–0,131	0,247	–0,148	0,000
18	3,438	0,036	0,035	0,959	–0,382	–0,111	0,050	–0,192	0,314
19	4,412	0,006	0,045	1,083	–0,160	0,117	–0,093	–0,002	–0,110
20	2,095	0,001	0,021	1,070	0,065	–0,036	0,053	0,001	–0,001
21	2,255	0,000	0,023	1,078	–0,002	–0,001	–0,001	0,001	0,001
22	5,908	0,004	0,060	1,111	0,126	0,062	0,020	–0,099	–0,034
23	1,971	0,000	0,020	1,074	–0,026	0,019	–0,013	–0,008	–0,013
24	2,432	0,002	0,025	1,071	0,083	0,063	–0,035	–0,057	0,003
25	0,596	0,002	0,006	1,043	–0,078	0,029	–0,033	–0,031	0,011
26	0,417	0,000	0,004	1,055	–0,027	0,004	0,001	–0,014	0,000
27	2,548	0,003	0,026	1,069	0,100	0,077	–0,045	–0,067	0,004
28	4,425	0,003	0,045	1,092	–0,118	0,083	–0,069	0,006	–0,082
29	3,748	0,010	0,038	1,058	–0,205	0,052	–0,128	–0,043	0,120
30	4,055	0,001	0,041	1,094	–0,072	–0,018	0,049	–0,044	–0,001
31	2,079	0,001	0,021	1,071	0,058	0,030	–0,046	0,008	0,011
32	1,132	0,000	0,011	1,066	–0,006	0,004	–0,003	–0,001	–0,003
33	2,304	0,006	0,023	1,049	0,152	–0,075	0,077	0,091	–0,047
34	2,421	0,000	0,024	1,079	0,020	0,003	–0,012	0,007	0,008
35	5,491	0,001	0,055	1,113	0,063	0,042	–0,003	–0,029	–0,042
36	0,421	0,001	0,004	1,046	–0,061	–0,036	0,013	0,024	0,012
37	3,912	0,000	0,040	1,096	–0,041	–0,018	0,030	–0,023	0,010
38	1,331	0,000	0,013	1,068	–0,005	0,001	–0,003	0,003	0,000
39	6,388	0,000	0,065	1,127	0,013	0,008	–0,002	–0,011	0,002
40	2,359	0,018	0,024	0,990	–0,269	–0,238	0,151	0,131	0,060
41	0,227	0,001	0,002	1,042	–0,063	–0,031	0,011	0,021	0,007
42	8,040	0,000	0,081	1,147	–0,020	0,014	–0,009	0,000	–0,017
43	5,755	0,001	0,058	1,116	–0,072	0,007	–0,024	–0,036	0,056
44	0,370	0,001	0,004	1,051	–0,045	0,010	–0,018	–0,010	0,010
45	2,768	0,002	0,028	1,075	0,087	0,075	–0,046	–0,014	–0,053
46	5,735	0,002	0,058	1,114	–0,088	0,014	–0,031	–0,047	0,064
47	3,393	0,007	0,034	1,062	0,172	–0,094	0,131	–0,060	0,074
48	4,431	0,000	0,045	1,103	–0,003	–0,002	0,002	–0,002	0,001
49	0,553	0,012	0,006	0,932	0,221	–0,048	0,112	0,027	–0,062
50	5,327	0,002	0,054	1,107	0,097	0,011	–0,004	0,062	–0,076
51	1,270	0,000	0,013	1,067	0,011	0,003	0,002	–0,008	0,000

(Continúa)

TABLA 4A.4. (Continuación)

Observación	Distancia de Mahalanobis	Distancia de Cook	Apalancamiento (valores sombrero)	COVRATIO	SDFFIT	SDFBETA			
						Constante	X_3	X_5	X_6
52	1,047	0,000	0,011	1,063	0,029	−0,006	0,003	0,020	−0,009
53	5,231	0,005	0,053	1,099	−0,139	0,004	0,070	−0,098	−0,026
54	0,989	0,012	0,010	0,960	−0,225	−0,166	0,132	0,078	−0,001
55	3,639	0,001	0,037	1,091	0,052	0,016	0,016	−0,040	−0,007
56	1,687	0,000	0,017	1,069	0,043	−0,005	−,002	−0,012	0,034
57	6,730	0,010	0,068	1,110	0,196	−0,081	−0,031	0,135	0,085
58	3,397	0,012	0,034	1,044	0,221	−0,161	0,121	0,136	0,016
59	0,199	0,000	0,002	1,055	−0,007	−0,002	0,002	−0,002	0,002
60	0,821	0,000	0,008	1,058	0,040	0,004	0,000	0,020	−0,022
61	2,945	0,005	0,030	1,065	0,138	−0,072	0,105	−0,048	0,053
62	2,056	0,010	0,021	1,022	0,198	−0,084	0,096	0,112	−0,076
63	1,473	0,000	0,015	1,068	−0,027	0,008	−0,018	0,011	−0,004
64	3,676	0,021	0,037	1,018	−0,292	−0,168	0,242	−0,111	0,046
65	4,496	0,025	0,045	1,029	0,314	0,121	−0,044	2,261	0,152
66	0,978	0,001	0,010	1,055	−0,062	0,005	−0,036	0,017	0,015
67	3,565	0,010	0,036	1,057	0,196	−0,131	0,091	0,005	0,144
68	1,288	0,006	0,013	1,020	0,160	0,106	−0,117	0,027	−0,032
69	2,186	0,019	0,022	0,980	0,275	−0,014	0,050	0,157	−0,184
70	0,346	0,000	0,003	1,052	0,037	0,002	0,008	0,007	−0,016
71	7,034	0,004	0,071	1,126	−0,124	−0,014	0,070	−0,097	0,013
72	5,257	0,001	0,053	1,110	0,067	−0,001	0,001	0,049	−0,048
73	2,996	0,000	0,030	1,085	−0,032	0,009	−0,022	−0,002	0,015
74	0,964	0,012	0,010	0,958	0,225	−0,119	0,135	0,072	−0,006
75	1,292	0,003	0,013	1,044	0,111	0,080	−0,065	−0,048	0,013
76	1,274	0,006	0,013	1,022	−0,155	0,088	−0,109	−0,030	−0,001
77	1,064	0,000	0,011	1,065	0,003	0,001	0,001	−0,002	0,000
78	3,916	0,008	0,040	1,068	−0,183	0,020	−0,101	−0,021	0,123
79	4,808	0,028	0,049	1,026	0,339	0,149	−0,055	0,288	0,144
80	1,001	0,012	0,010	0,967	0,218	−0,016	0,081	0,061	−0,122
81	2,014	0,014	0,020	0,998	0,236	−0,049	0,064	0,147	−0,130
82	8,911	0,005	0,090	1,150	−0,137	0,000	0,075	−0,046	−0,081
83	4,964	0,015	0,050	1,068	0,241	0,083	−0,114	−0,110	0,164
84	0,129	0,011	0,001	0,897	−0,212	−0,068	0,047	−0,037	0,052
85	0,284	0,005	0,003	0,987	0,146	−0,027	0,059	−0,031	0,031
86	0,809	0,001	0,008	1,057	0,045	0,031	−0,023	0,001	−0,021
87	0,250	0,004	0,003	1,004	0,124	−0,006	0,043	−0,037	0,015
88	0,363	0,009	0,004	0,947	−0,191	−0,108	0,061	0,002	0,077
89	1,612	0,003	0,016	1,051	−0,109	−0,068	0,023	0,079	−0,004
90	4,176	0,010	0,042	1,068	0,197	−0,137	0,092	0,013	0,145
91	2,230	0,000	0,023	1,076	0,041	−0,010	0,000	−0,006	0,034
92	3,682	0,035	0,037	0,970	−0,380	−0,135	0,060	−0,168	0,322
93	7,266	0,001	0,073	1,136	−0,054	−0,012	0,038	−0,012	−0,027
94	5,030	0,026	0,051	1,039	0,321	0,285	−0,266	−0,101	−0,012
95	1,862	0,009	0,019	1,019	−0,192	−0,165	0,094	0,083	0,064
96	9,485	0,005	0,096	1,159	0,134	0,098	−0,037	−0,115	−0,003
97	1,099	0,012	0,011	0,973	0,217	−0,127	0,137	0,062	0,015
98	4,101	0,018	0,041	1,041	0,266	0,219	−0,223	−0,067	0,014
99	1,015	0,001	0,010	1,055	0,066	0,039	−0,038	−0,021	0,018
100	2,186	0,062	0,022	0,778	−0,517	0,084	−0,261	0,335	−0,212
Mínimo	0,129	0,000	0,001	0,771	−0,937	−0,238	−0,402	−0,288	−0,759
Máximo	9,485	0,201	0,096	1,159	0,339	0,657	0,247	0,341	0,322

Nota: Los valores que exceden de los umbrales están subrayados.

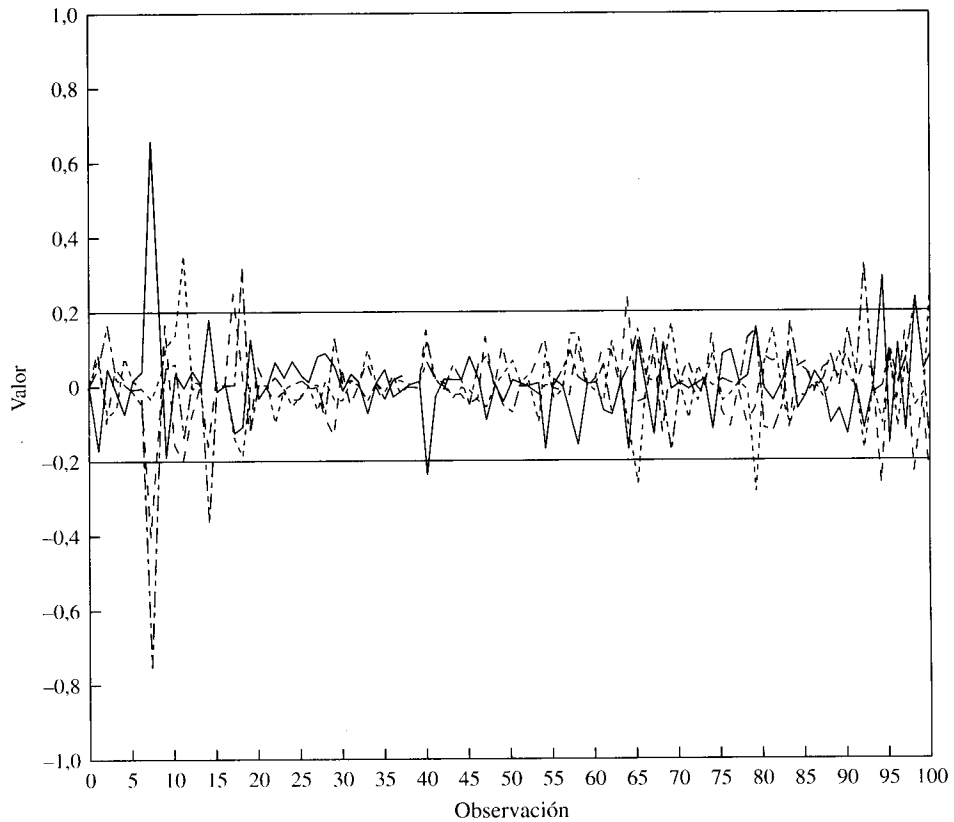


FIGURA 4A.4. SDFBETA para la constante X_3, X_5 y X_6.

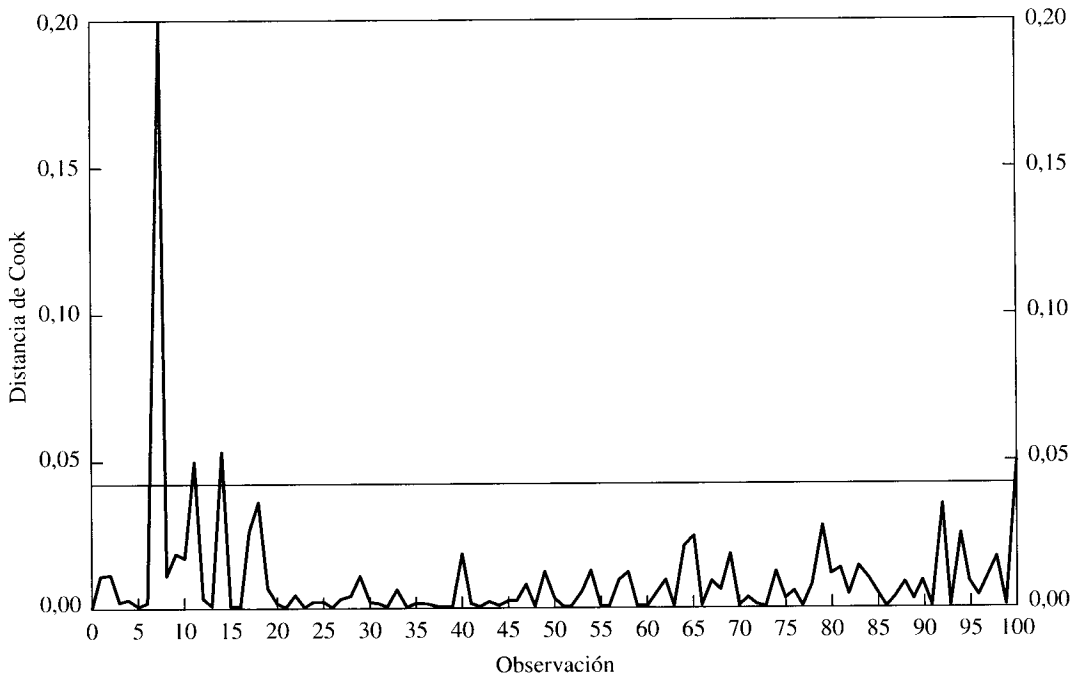


FIGURA 4A.5. Distancia de Cook.

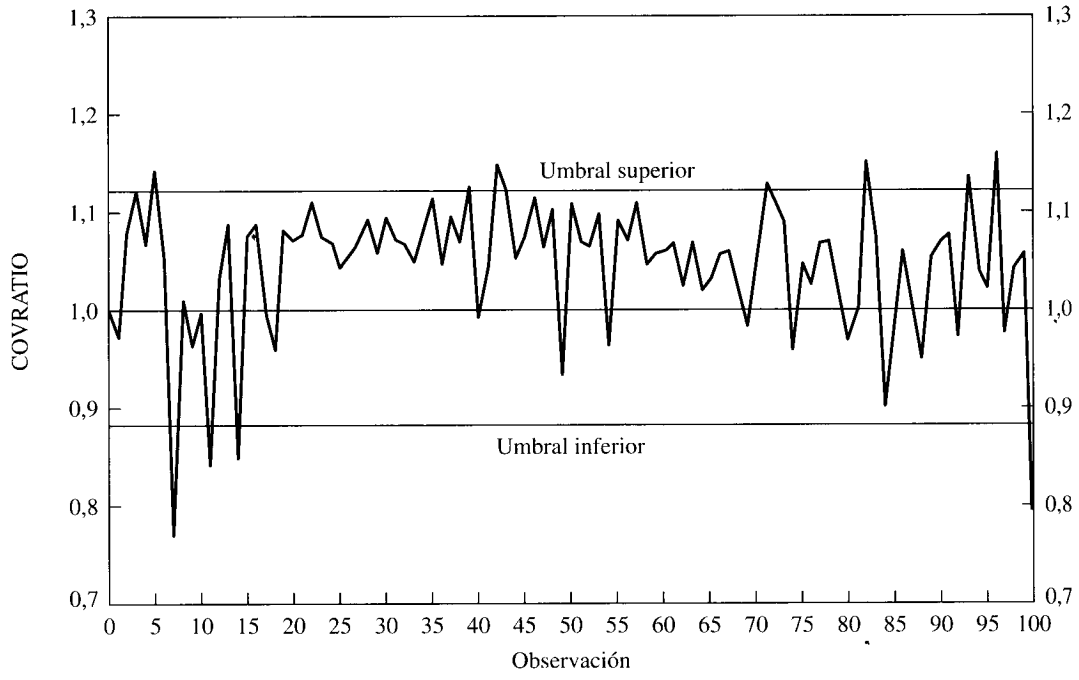


FIGURA 4A.6. COVRATIO.

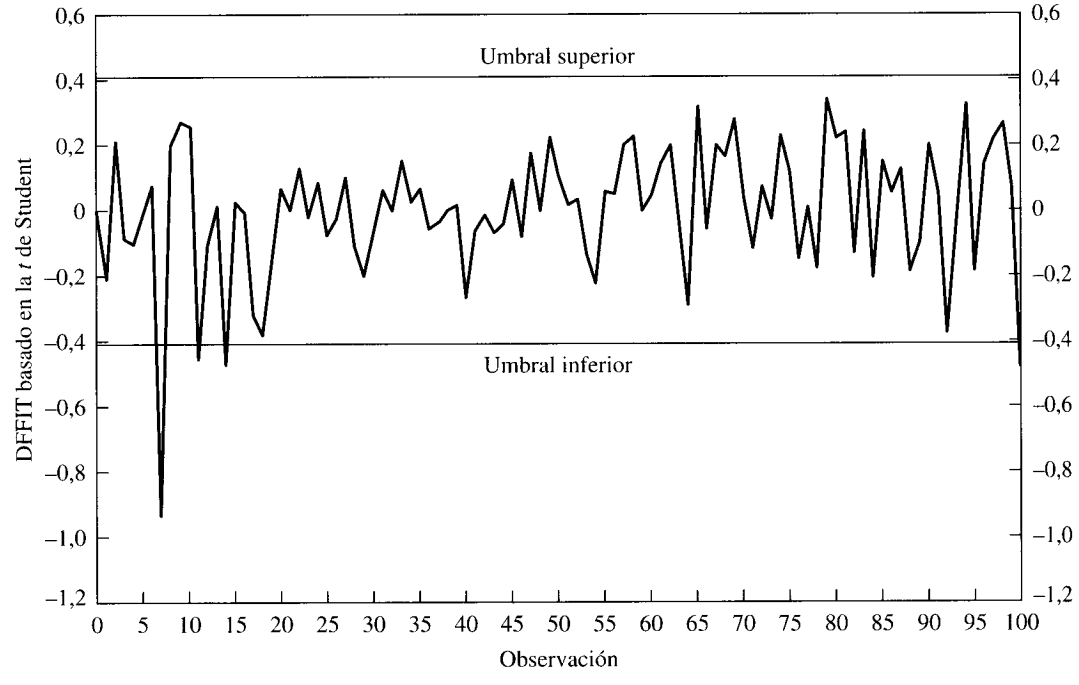


FIGURA 4A.7. DFFIT estandarizada (SDFFIT).

las otras observaciones en casi todas de las medidas diagnósticas, pero también se debe considerar las otras tres observaciones como influyentes dada su identificación consistente en casi todas las medidas.

TABLA 4A.5. Resumen de las pruebas diagnósticas para las observaciones influyentes

Medida de diagnosis	*Especificación del valor umbral*	*Valor umbral calculado*	*Observaciones que exceden el umbral*[a]
Residuos			
Estandarizados	Valor t crítico a un nivel de confianza especificado	±1,96	**7,** 11, 14, 100
Basados en la t de Student	Valor t crítico a un nivel de confianza especificado	±1,96	**7,** 11, 14, 100
Basados en la t de Student eliminados	Valor t crítico a un nivel de confianza especificado	±1,96	**7,** 11, 14, 84, 100
Apalancamiento			
Valores sombrero	Muestra pequeña: $3(k11)/n$	0,12	Ninguna
	Muestra mediana/grande: $2(k11)/n$	0,08	5, 42, 82, **96**
Distancia de Mahalanobis	Evaluar la distribución de los valores	Ninguna	Las diez primeras observaciones: **96,** 82, 5, 42, 93, 7, 71, 57, 3, 39
Medidas de caso único			
SDFBETA	Muestra pequeña: valor t crítico a un nivel de confianza especificado	±1,96	Ninguna
			Constante: **7,** 40, 94, 98
		0,02	X_3: **7,** 17, 40, 64, 94, 98, 100
	Muestra mediana/grande: $2/\sqrt{n}$		X_5: **11,** 65, 79, 100
			X_6: **7,** 14, 18, 92, 100
Distancia de Cook	$\frac{4}{n-k-1}$	0,042	**7,** 11, 14, 100
COVRATIO	$1 \pm \frac{3(k+1)}{n}$	Superior: 1,12	Superior: 3, 5, 39, 42, 71, 82, 93, **96**
		Inferior: 0,88	Inferior: **7,** 11, 14, 100
SDFFIT	$2\sqrt{\frac{k!\ \ 1}{n.\ \ k.\ \ 1}}$	0,408	**7,** 11, 14, 100

[a] Observaciones con valores máximos aparecen en **negrita**.
n = Tamaño muestral.
k = número de variables independientes.

Cuarto paso: *Selección y tratamiento de los casos influyentes*

Aunque no existe un procedimiento único para identificar los casos influyentes y decidir después un curso de acción, la premisa básica es muy simple. En ausencia de errores de entrada de datos o de otras medidas de corrección, deberían examinarse atentamente los casos influyentes que son sustancialmente diferentes del resto de los datos de una o varias variables.

Si se llega a la conclusión de que un caso no es representativo de la población general, debería eliminarse. Nuestro objetivo es estimar la ecuación de regresión sobre una muestra representativa para obtener resultados generalizables. Si la muestra contiene una o más observaciones que no son representativas, nos impide conseguir nuestro objetivo.

A efectos de la exposición, seleccionaremos cuatro casos (7, 11, 14 y 100) para su eliminación. Estos casos fueron consistentemente identificados por los análisis de diagnóstico y se consideran los casos con el mayor impacto sobre la mejora de la ecuación. La Tabla 4A.6 muestra el modelo de regresión final con estos cuatro casos eliminados. Comparando estos resultados con aquéllos en el Capítulo 4, observamos sustanciales mejoras en todos los aspectos. Se mejora la predicción conjunta, con un R^2 que cambia de 0,768 a 0,833, además del efecto que obtenemos de la adición de

TABLA 4A.6. Resultados de la regresión múltiple después de eliminar cuatro observaciones influyentes

Resultados del modelo de regresión conjunta

R Múltiple	0,913
R^2 Múltiple	0,833
R^2 Ajustado	0,828
Error estándar de la estimación	3,698

Análisis de la varianza

	Suma de los cuadrados	*Grados de libertad*	*Cuadrado media*	*Ratio F*
Regresión	6291,8	3	2097,3	153,37
Residuos	1258,1	92	13,7	

Variables en la ecuación

Variables	*Coeficiente*	*Error estándar del coeficiente*	*Coeficiente de regresión estandarizado*	*Valor t parcial*
Constante	–9,645	2,803		–3,440
X_3 Flexibilidad de precio	3,719	0,274	0,582	13,566
X_5 Servicio global	7,094	0,521	0,601	13,605
X_6 Imagen del personal de ventas	2,337	0,521	0,198	4,482

Variables no presentes en la ecuación

	Correlación parcial	*Valor t*
X_1 Velocidad de entrega	–0,004	–0,036
X_2 Nivel de precios	–0,019	–0,177
X_4 Imagen del fabricante	–0,116	–1,117
X_7 Calidad de producto	0,136	1,311

la tercera variable (X_6) a la ecuación. También, el error estándar disminuye de 4,39 a 3,69, una mejora del 16 por ciento. Además, cada coeficiente mejora en significación estadística, indicando un fortalecimiento de las relaciones al eliminar estos casos atípicos influyentes.

Visión panorámica

La identificación de los casos influyentes en un paso esencial en la interpretación de los resultados del análisis de regresión. El investigador debe tener cuidado, sin embargo, en la eliminación discrecional de los casos considerados como influyentes. Siempre existen casos atípicos en cualquier población y se debe recortar el conjunto de datos de tal forma que se garanticen siempre unos buenos resultados. No obstante, se debe intentar conseguir la mejor representación de la relación de la muestra, y la influencia de unos pocos casos puede distorsionar o inhibir completamente la consecución de este objetivo. Por tanto, recomendamos que se utilicen estas técnicas con conocimiento y precaución, en la medida en que pueden representar tanto beneficios como daños potenciales.

Resumen

En la medida en que el análisis de regresión aumenta tanto en alcance como en complejidad, se hace esencial explorar los asuntos de este apéndice. Tanto las observaciones influyentes como la multicolinealidad pueden tener un impacto sustancial sobre los resultados y su interpretación. Sin embargo, avances recientes en las técnicas de diagnóstico como las que se han expuesto anteriormente proporcionan al investigador un método simplificado de realización de los análisis que identificarán los problemas en estas áreas. Dondequiera que el análisis de regresión encuentre una de estas áreas de problemas, se anima al investigador a investigar las cuestiones expuestas aquí y adoptar las soluciones adecuadas si fuera necesario.

Cuestiones de revisión

1. Describa las razones para no depender solamente en la matriz de correlación univariante para el diagnóstico de la multicolinealidad.
2. ¿En qué casos la detección de casos atípicos potencialmente pierde otras observaciones influyentes?
3. Describa la diferencia en el uso de los residuos (incluyendo los residuos Basados en la *t* de Student) frente a los diagnósticos de caso único de DFBETA y DFFIT.
4. ¿Qué criterios sugeriría para determinar si una observación se tiene que eliminar del análisis?

Bibliografía

1. Barnett, V., and T. Lewis (1984), *Outliers in Statistical Data*, 2d ed. New York: Wiley.
2. Belsley, D. A., E. Kuh, and R. E. Welsch (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
3. Rousseeuw, P. J., and A. M. Leroy (1987), *Robust Regression and Outlier Detection*. New York: Wiley.
4. Weisberg, S. (1985), *Applied Linear Regression*. New York: Wiley.

LECTURA COMPLEMENTARIA

Relación entre concentración y precios en el comercio minorista

M.ª JESÚS YAGÜE GUILLÉN
Universidad Autónoma de Madrid

Información Comercial Española ICE, n.º 739
Marzo 1995

En el presente artículo se trata la relación entre tamaño, concentración y poder de mercado aplicada al ámbito de la distribución minorista. Para ello, se exponen algunos rasgos diferenciales característicos de los mercados de distribución en general, así como algunas medidas que pretenden aproximar la concentración del sector en nuestro país. El trabajo se completa con un análisis empírico en el que se estudia la relación entre concentración y precios en el comercio de alimentación español.

Palabras clave: *comercio interior comercio minorista, alimentos, concentración industrial, marcas, precios, competencia, modelo econométrico, España.*

1. IMPORTANCIA DEL ANÁLISIS Y MEDICIÓN DEL GRADO DE CONCENTRACIÓN DE LOS MERCADOS

Las proposiciones sobre la eficiencia de los sistemas competitivos, procedentes de la Teoría del Bienestar[1], aseguran que el sistema de precios garantiza que la asignación alcanzada en la economía es óptima en el sentido de Pareto. En mercados de competencia perfecta, los consumidores obtienen los productos deseados en las mejores condiciones de precios posibles y las empresas alcanzan beneficios normales, ya que la actuación de la libertad de entrada y salida en los mercados, conduce a que a largo plazo desaparezcan los beneficios extraordinarios. Por el contrario, en mercados de competencia imperfecta, modelo más realista, las proposiciones teóricas muestran que existen incentivos, en términos de mayores beneficios conjuntos, para que las empresas lleven a cabo procesos de concentración económica y/o estrategias cooperativas con otras empresas que las dirijan hacia posiciones de dominio en los mercados. Puesto que la cooperación entre empresas genera mayores precios, mayores beneficios y menores cantidades producidas, se estima que los intereses de los consumidores no se verán satisfechos en mercados fuertemente concentrados.

Estos resultados teóricos sugieren que si hay pocas empresas en el mercado y no existe la presión de la entrada de nuevos competidores, las instaladas obtendrán unos márgenes precio coste marginal positivos, garantizando la consecución de beneficios supracompetitivos[2]. Por tanto, las empresas desarrollarán estrategias tratando de aprovechar las oportunidades que les brinda la estructura de competencia propia de su mercado, buscando formas de poder más concentrado y en última instancia de monopolio.

No obstante, en los modelos teóricos se reconoce que ciertos grados de imperfección y en el extremo el monopolio, son justificables, por ejemplo, en base al aprovechamiento de economías de costes: escala, alcance o aprendizaje[3]. Es por esta razón, por lo que no existe a *priori* una presunción de que las posiciones de dominio sean necesariamente malas, sino que lo no deseable es el abuso de una situación dominante[4].

[1] Véanse los trabajos de ARROW (1951), DEBREU (1959) y KOOPMANS (1957) para encontrar las proposiciones y las pruebas más importantes de la economía del bienestar.

[2] Este resultado puede no ser cierto si las empresas guían sus actuaciones por objetivos distintos de la maximización del beneficio, como por ejemplo la maximización de la penetración del mercado.

[3] Existen otras situaciones que pueden justificar la existencia de poder de monopolio. El caso de las patentes constituye una buena ilustración. A corto plazo puede ser eficiente que la información sobre ciertas mejoras tecnológicas se difunda rápidamente, pero por otro lado si no se garantiza una recuperación de las inversiones, las empresas no tienen incentivos para innovar.

[4] A este respecto, es necesario tomar en consideración las propuestas derivadas de la literatura de mercados contestables.

En consecuencia, dada la importancia que los modelos teóricos conceden a la estructura competitiva de los mercados, y por ende, al grado de concentración, para explicar los niveles de beneficios, de eficiencia y de bienestar alcanzables en los mismos, es crucial para la toma de decisiones de las empresas y de las autoridades públicas, disponer de instrumentos analíticos y empíricos que permitan transformar los conceptos abstractos que caracterizan a los mercados, en categorías operativas precisas, capaces de ser utilizadas para determinar o medir, mediante indicadores objetivos, la intensidad de la rivalidad en los mismos.

Cuando se trata de disponer de herramientas operativas útiles para analizar las estructuras de competencia de los mercados, es preciso recurrir a las aportaciones procedentes de la Economía Industrial y derivadas de su paradigma básico estructura-conducta-resultados[5]. La mayoría de los trabajos teóricos y empíricos, desarrollados en torno a este paradigma han tenido por objeto el estudio de los mercados industriales. Sólo recientemente, desde otros campos de la investigación económica, se están aplicando el paradigma y los resultados procedentes de la Economía Industrial a mercados de servicios[6].

Las dos posturas radicalmente distintas, estructuralista y conductista, aparecidas en la interpretación del paradigma, otorgan un papel protagonista a la concentración, como variable determinante, en un caso, o *proxy*, en el otro, de los resultados. Según la primera, en los mercados más concentrados se alcanzan tasas de beneficios significativamente superiores que en los mercados menos concentrados. La validación empírica de esta hipótesis desde los trabajos pioneros de Bain (1951,1956 y 1959) ha consistido en establecer una asociación positiva y estadísticamente significativa, principalmente sobre bases de datos transversales, entre algún indicador del nivel de concentración de los mercados y sus márgenes medios de beneficio[7]. Aunque la mayoría de los trabajos han confirmado la hipótesis de partida, la valoración de la relación encontrada ha dado lugar a un profundo debate que se centra en los siguientes aspectos:

a) Cuando los precios de los bienes son observables y por ello directamente comparables, el estudio de la relación concentración-poder de mercado debe realizarse sobre dichos datos.

b) Cuando la relación debe estudiarse mediante márgenes de beneficio, surgen dos interpretaciones contradictorias sobre el signo positivo, la de ejercicio de poder de monopolio y la de obtención de mayor eficiencia.

c) En muchos estudios, la relación encontrada es estadísticamente débil y la magnitud del coeficiente es baja.

d) Se hace necesario el estudio de la relación incluyendo factores interactivos que reconozcan que el efecto de la concentración sobre los beneficios de monopolio sólo puede perdurar si a la vez existen o se generan barreras a la entrada.

El análisis teórico y la evidencia empírica no son en absoluto concluyentes sobre si la presencia de empresas de gran tamaño en los mercados, y la elevada concentración que pueden suponer tienen efectos netamente positivos o negativos sobre su grado de eficiencia estática y/o dinámica[8].

En este trabajo no pretendemos aportar elementos que cierren dicha discusión. Por el contrario, se trata de plantear el problema de la relación entre tamaño, concentración y poder de mercado en un contexto diferente al de los mercados industriales, en concreto en los mercados de distribución minorista, donde las dificultades de tratamiento conceptual y empírico son todavía más acusadas.

En el epígrafe segundo abordaremos algunos rasgos diferenciales que caracterizan a los mercados de distribución. En el tercero, plantearemos algunas medidas que pretenden aproximar la concentración actual en el sector de la distribución en España; y en el cuarto y último presentamos un análisis empírico en el que se estudia la relación entre concentración y precios en el comercio de alimentación en nuestro país.

2. LOS MERCADOS DE DISTRIBUCIÓN MINORISTA

El estudio del grado de competencia en los mercados de distribución no puede realizarse mediante la mera traslación de los argumentos teóricos y de los instrumentos empíricos tradicionalmente aplicados a los mercados industriales. Aunque son múltiples los rasgos diferenciales entre ambos tipos de mercado, aquí sólo presentaremos aquellos que, en nuestra opinión, son más relevantes desde la perspectiva del análisis concurrencial.

En primer lugar, es preciso comenzar definiendo la actividad y el producto de las empresas minoristas. Aunque durante mucho tiempo este tipo de empresas ha sido concebido como un mero intermediario entre fabricantes y consumidores, con escasa capacidad decisoria, puesta en la mayor parte de las ocasiones al servicio de los objetivos e intereses de los fabricantes, en la actualidad hay unanimidad entre los autores al definir la actividad minorista como la prestación de un conjunto de servicios al consumidor que tienen por objeto reducir los costes en los que éste debe incurrir para transformar los productos elaborados por las empresas de fabricación en utilidades o satisfacción. Si-

[5] Aunque la literatura de Dirección Estratégica ha aportado un marco conceptual para el análisis de la rivalidad en los mercados, el «Núcleo Competitivo de Porter», en nuestra opinión este nuevo modelo de análisis debe entenderse más como una reformulación y extensión del paradigma de la Economía Industrial que coma un modelo radicalmente diferente.

[6] Algunas traboos representativos de la aplicación de la Economía Industrial a los servicios de comercio minorista son: NELSON (19911, PERRY (1991), SHAFFER (1991), NOOTEBOOM (1985), CARREE y THURIK (1990), STEINER (1993).

[7] Para una revisión puede consultarse SCHERER (1980), SCHERER y ROSS (1990) y SCHMALENSEE (1989).

[8] Los dos primeros capítulos del libro «Concentración empresarial y competitividad: España en la CEE» recogen una magnífica exposición de los argumentos teóricos y empíricos que apoyan el que la polémica sobre la conveniencia o inconveniencia de las empresas de gran tamaño en los mercados siga abierta.

guiendo a Betancourt y Gautschi (1988) la empresa minorista proporciona bienes y servicios ofrecidos para la compra a un precio explícito junto a un conjunto de servicios de distribución (localización, información, ambiente, conveniencia, surtido) que no son explícitamente vendidos en el mercado. El conjunto de servicios ofrecido es el que sirve para caracterizar las diferentes organizaciones de venta minorista existentes, sus estrategias de segmentación y posicionamiento y en consecuencia, su forma de competir en el mercado.

En segundo lugar, la posición detentada por las empresas minoristas entre los fabricantes-mayoristas y los consumidores los convierte en agentes económicos que actúan en los dos mercados, el industrial y el de consumo. La visión tradicional concibe a la empresa minorista como cliente de los mercados industriales, donde adquiere productos ofrecidos por los fabricantes- mayoristas a cambio de dinero y como proveedor en los mercados finales, donde vende productos y servicios explícitos e implícitos a cambio de dinero. En este contexto de doble etapa, el poder de mercado y los márgenes de un fabricante o de un minorista individual son una función conjunta de su posición competitiva horizontal frente a empresas del mismo nivel y de su poder negociador frente a las empresas del otro nivel.

Sin embargo, una moderna concepción del comercio minorista, Mulhern y Leone (1991) entiende que en los mercados industriales las empresas minoristas desempeñan una doble función, la tradicional de clientes, y la más actual de proveedores de segmentos de demanda o de consumidores, a cambio de lo cual esperan recibir un conjunto de condiciones privilegiadas en los componentes explícitos o implícitos de los contratos de compra de los productos. Por esta razón, en la relación con los fabricantes el estudio de la competencia vertical exige ampliar los modelos convencionales de integración vertical, en la medida en que los mercados de consumidores relevantes para los fabricantes están bajo el control de los minoristas.

Entre los aspectos concretos relacionados con el estudio del grado de competencia de los mercados minoristas en los que influyen los argumentos teóricos anteriores destacan los siguientes:

a) El producto ofrecido por las empresas minoristas, aun dentro del mismo sector o subsector comercial, es un producto intangible, diferenciado y de difícil medición. No obstante, la forma comercial (por ejemplo en alimentación, hipermercados, supermercados, tiendas de descuento, tradicionales, etc.), suele definir el producto o servicio global y el segmento de mercado al que se dirige el establecimiento. Por tanto, dentro de la competencia horizontal establecida en un sector comercial, debe tenerse en cuenta que hay un componente de competencia intertipo o competencia en mercados con productos diferenciados, que suele aproximarse a modelos de competencia monopolística, y otro componente de competencia más directa, intratipo o competencia entre empresas de productos homogéneos o cuasihomogéneos. El consumidor tipo de los hipermercados es bien diferente del de las tiendas de descuento o del de las tiendas tradicionales, y las características y la evolución de sus respectivos segmentos también lo son. Por tanto, en ausencia de regulación, la interdependencia de las demandas dirigidas a cada tipo de establecimiento es pequeña. Además, la actuación regulatoria podría llevar asociada grandes pérdidas de satisfacción (bienestar) de los consumidores, en la medida en que les obligue a acomodar o dirigir su demanda a tipos de establecimiento que ofrecen un producto poco adaptado a sus necesidades. Este doble componente de la competencia horizontal obliga a estudiar el grado dé concentración del mercado entre tipos de establecimientos y dentro de cada tipo.

b) La consideración del surtido como un componente del servicio global ofrecido por los comercios minoristas nos lleva a realizar las siguientes precisiones desde la perspectiva de la competencia. Cuando se trata con productos que habitualmente se compran como parte de una cesta (alimentación, por ejemplo) es necesario tener en cuenta que la estructura de competencia, horizontal y vertical, es muy diferente por familias, categorías o tipos de productos. En el Cuadro 4.L.1 se recogen los datos de la cuota de mercado que ostentan distintos tipos de establecimientos de alimentación por familias de productos. Su lectura indica una diferencia clara en la posición que en el mercado mantiene cada tipo de establecimiento en las ventas de productos frescos y secos. Mientras las tiendas tradicionales ostentan una cuota elevada en productos frescos, los hipermercados alcanzan las cuotas más altas en productos envasados como el aceite, la leche o el vino.

Otro aspecto del surtido a tener en cuenta es el referente a la marca. Dentro del surtido de las empresas comerciales pueden concurrir tres tipos de marcas, líderes, marcas de fabricante no líderes y marcas de distribuidor. la competencia dentro de cada tipo y entre tipos de marcas es diferente.

En las marcas líderes, la competencia intramarca es, en términos relativos, baja, mientras que la competencia intramarca entre establecimientos es relativamente alta. Este tipo de marcas se utiliza por los comerciantes para formar su imagen frente al consumidor, y ofrecer información sobre precios que el consumidor pueda utilizar en el análisis y comparación de establecimientos competidores y le ayude a realizar su elección. En las marcas líderes, los márgenes comerciales suelen ser bajos y los de los fabricantes altos, a la vez que se espera que la dispersión de precios entre los establecimientos sea relativamente baja[9]. En las marcas de fabricante desconocidas y en las marcas de distribuidor, la competencia entre establecimientos es baja o inexistente, mientras que es alta entre marcas en el interior del establecimiento comercial. En ellas, los márgenes comerciales suelen ser altos y los de los fabricantes bajos, además la dispersión de precios es superior.

La distribución de marcas líderes y no líderes por familias de productos no es homogénea. Así por ejemplo, mientras que en derivados cárnicos, conservas vegetales o productos congelados la cuota detentada por las marcas líderes no es muy elevada, en otros productos como sopas,

[9] En DEIKE (1992) se aprecia para algunas categorías de productos de alimentación esta evidencia.

CUADRO 4.L.1. Cuota de mercado de las formas comerciales en la demanda de alimentos de los hogares

Productos	*Tienda tradicional*		*Autoservicios*		*Hipermercados*		*Otros*	
	1987	*1992*	*1987*	*1992*	*1987*	*1992*	*1987*	*1992*
Huevos	48,48	36,04	28,65	30,90	1,88	4,67	20,99	28,39
Carnes	64,80	53,97	22,82	31,57	2,44	6,69	9,94	7,77
Pesca	76,72	67,81	13,77	21,51	1,73	6,89	7,78	3,79
Leche	30,44	17,95	39,97	51,23	4,62	16,13	24,97	14,67
Lácteos	37,53	25,85	47,76	54,09	5,34	14,08	9,37	5,98
Pan	79,45	78,09	13,52	15,74	1,02	2,35	6,01	3,82
Bollería	56,14	45,37	31,73	40,32	4,26	10,52	7,87	3,79
Arroz	34,45	18,92	52,90	60,93	5,15	17,69	7,50	2,46
Pastas	32,69	17,52	54,00	62,39	6,10	17,15	7,21	2,94
Legumbre	34,51	24,65	43,45	50,82	4,44	11,97	17,60	12,56
Aceites	24,70	14,31	50,51	52,65	6,89	18,36	17,90	14,68
Patatas	45,01	37,81	23,08	32,15	2,06	8,05	29,85	21,99
Hortalizas	56,00	49,48	21,34	23,70	1,64	3,55	21,02	23,27
Frutas	59,82	53,59	23,03	25,85	1,68	3,94	15,47	16,62
C. vegetal	27,68	19,53	55,63	57,42	7,34	17,38	9,35	5,67
Vinos (*)	28,37	13,22	37,46	51,70	10,61	21,70	23,56	13,38
Vinos (**)	39,98	32,96	24,77	35,58	2,03	9,54	33,22	21,92
Cerveza	42,08	22,58	36,36	50,17	4,76	14,89	16,80	12,36
Zumo	37,77	21,90	49,32	59,72	5,17	15,14	7,74	3,4
TOTAL	53,28	46,13	29,94	35,04	3,46	8,93	13,32	9,90

(*): Vino VCPRD; (**): Vino de mesa.
Fuente: *La alimentación en España*, 1992, MAPA.

café instantáneo, derivados lácteos o refrescos la dominancia de las marcas líderes es mucho mayor.

Uno de los aspectos que afectan de forma más notable a la competencia vertical consiste en la presencia y extensión de las marcas de distribuidor. Aunque en España entraron más tarde que en el resto de Europa y no han experimentado un crecimiento tan rápido, se van consolidando, y en 1993 representaban una cuota del 10 por ciento en alimentación, porcentaje todavía alejado del de países como el Reino Unido con el 27,1 por ciento o el de Bélgica con el 21,7. Las marcas de distribuidor nacen y se desarrollan de la mano del crecimiento de los grandes formatos de distribución. El aumento de la concentración de las ventas en hipermercados y grandes cadenas de supermercados o de tiendas de descuento favorece el aumento de la cuota de marcas de distribuidor, las cuales representan productos más baratos y presionan la competencia entre marcas de fabricante con escasa notoriedad y marcas de distribuidor, pero también entre estas últimas y marcas lideres. En las categorías de productos donde hay más marcas de distribuidor, los márgenes comerciales son mayores y los de fabricante son menores, los precios se mantienen más bajos y menos dispersos.La asociación entre importancia de la marca de distribuidor y la competencia en precios también se ha visto favorecida porque desde su origen las marcas de distribución han encontrado un caldo de cultivo más propicio en las categorías de productos poco diferenciadas o innovadoras: tomate natural, legumbres precocidas, o mermeladas[10].

c) Los servicios de localización, acceso y conveniencia que intervienen en la creación del producto de la empresa minorista, fuerzan a introducir en el análisis de la competencia del sector el ámbito geográfico. De un lado, la competencia horizontal en los mercados de consumo final se establece a nivel local y por tanto la medición del grado de concentración y su impacto sobre los resultados debe tomar como unidades de análisis el establecimiento comercial y la ciudad, el barrio o incluso la calle. Por otro lado, la competencia vertical exige considerar el mercado relevante a nivel nacional, europeo o incluso global, ya que en los mercados industriales la forma de competir también lo es. Por tanto, los minoristas cuando actúan en dichos mercados, bien como clientes que deben negociar las condiciones de compra de los productos, bien como vendedores que ofrecen determinados segmentos de consumidores finales a las empresas fabricantes, compiten a nivel global frente a otros minoristas o frente a los fabricantes. En este contexto, las unidades de

[10] Según cifras de Dym Panel publicadas en el número 219 de Distribución Actualidad, las cuotas de marcas de distribuidor en los tres productos citados serían, respectivamente, del 31,28,2 y 25,9 por ciento.

análisis son las empresas, en las que se encuentren integrados o asociados los establecimientos y el mercado nacional o europeo. En consecuencia, mientras un componente del margen comercial y del precio final depende de la estructura de competencia y del grado de concentración existente en el mercado local, otro componente varía con la estructura de competencia y el grado de concentración existente en el mercado nacional o internacional.

d) En el análisis concurrencial los precios constituyen la variable que mejor refleja el grado de competencia del mercado. Sin embargo, la observabilidad, el conocimiento y la cuantificación de los precios en los mercados minoristas son prácticamente imposibles. En los mercados industriales, las empresas disponen de carteras de productos más bien reducidas en relación a las empresas minoristas, el precio de venta de cada referencia es una información perfectamente conocida por los clientes, se modifica pocas o ninguna vez a lo largo del año y puede ser utilizada en el análisis comparativo de empresas rivales. Por el contrario, en los mercados finales las empresas minoristas disponen de surtidos muy amplios de productos, el precio de cada referencia, como ya se ha dicho, tiene un componente explícito y otro implícito, además se modifica con mucha frecuencia a lo largo del año, y sólo se publicitan fuera del establecimiento comercial los precios de algunas de las referencias. Todas estas circunstancias, junto a la racionalidad limitada de los individuos, convierten a los precios de venta al público en una información escasa y mal conocida, tanto por los consumidores como por los competidores. Por tanto, los consumidores en su elección de establecimiento se guían más por precios relativos que por absolutos, por precios recordados que por reales, y en las compras agrupadas, los refieren a una pequeña cesta acorde con su estructura de consumo más permanente. Todo ello, sugiere la gran magnitud del problema que rodea la medición y la comparación de los precios practicados por las empresas minoristas, y en consecuencia la cuantificación de las pérdidas de bienestar que se derivarían de existir un ejercicio abusivo de una posición dominante en el mercado.

3. MEDICIÓN DE LA CONCENTRACIÓN EN EL SECTOR DE LA DISTRIBUCIÓN EN ESPAÑA

Una revisión de la literatura procedente del ámbito de la Economía Industrial, pronto advierte que el estudio de las medidas de concentración de los mercados siempre ha sido un tema de máximo interés, se han desarrollado baterías de indicadores alternativos, se han estudiado sus propiedades estadísticas, su relación con medidas de poder de monopolio bajo diferentes formas de competencia en el mercado, etc. Sin embargo, desde una perspectiva puramente empírica, la mayoría de los trabajos ha utilizado índices de concentración simples, en concreto los más ampliamente aplicados son el índice absoluto, cuota de mercado de las *n* mayores empresas del mercado y el índice de Herfindahl.

Algunos de los rasgos diferenciales de los mercados en los que compiten las empresas minoristas (competencia intramarca e intraestablecimiento, competencia intertipo, competencia global y local, competencia horizontal y vertical) limitan la validez de la aplicación de los indicadores tradicionales a estos mercados. Además, la pobreza de las fuentes estadísticas sectoriales en nuestro país impiden, no sólo la confección de medidas más apropiadas, sino también la correcta elaboración de los indicadores tradicionales.

En este apartado presentamos algunas medidas que, aunque débiles, ayudan a extraer ciertas conclusiones sobre la situación de la concentración en el comercio minorista español. Para elagregado del comercio, en el Cuadro 4.L.2 se presentan las estimaciones realizadas de las cuotas de mercado de las cuatro, cinco y 10 mayores empresas del sector comercial, utilizando las cifras de ventas de las principales empresas comerciales publicadas por fomento de la producción y las cifras sobre valor de la producción del sector comercio elaboradas a partir de la contabilidad nacional de España[11].

CUADRO 4.L.2. Índices de concentración agregados para el sector comercio

Cuotas de mercado	*1990*	*1991*	*1992*	*1993*
C_4	18,2	20,1	19,0	18,0
C_5	19,9	21,9	20,8	19,9
C_{10}	25,4	27,7	26,4	25,7

La lectura de este cuadro advierte dos hechos. El primero se refiere a que la participación de las mayores empresas en la actividad comercial total no es elevada, ya que sólo supera ligeramente el 25 por ciento del total, el segundo pone de manifiesto que desde 1991 el sector comercial reduce ligeramente su grado de concentración medido por este indicador. Este dato, probablemente se explica por el efecto que el descenso del consumo privado de estos años de crisis tiene sobre la cifra de negocios de los grandes almacenes: Galerías Preciados y el Corte Inglés.

El análisis sectorial de las empresas comerciales que ocupan las primeras posiciones en el *ranking* por cifra de negocios muestra que en 1993 entre las cuatro mayores, tres son grandes superficies. de alimentación y una es un gran almacén y entre las 10 primeras, siete son grandes superficies o cadenas de supermercados de alimentación y tres son grandes almacenes o cadenas de confección. Lo cual indica que la concentración comercial se proyecta sobre estos dos subsectores, ya que si excluimos los de carburantes y tabaco, por estar en gran medida o totalmente integrados hacia arriba, el resto de los sectores comerciales no se caracteriza por estar dominado por grandes empresas. Esta

[11] Se ha tomado la participación que en el valor añadido bruto (VAB) de los servicios destinados a la venta tuvo el sector comercio en 1990 y se ha aplicado a las cifras sobre VAB de los servicios destinados a la venta de los años siguientes. Además, se ha tomado la participación que los consumos intermedios tuvieron en el valor de la producción del sector comercio en 1989 y se ha añadido a las cantidades anteriormente calculadas.

circunstancia, sin embargo, no significa que en determinados mercados locales, o para determinadas categorías de productos, no pueda haber una situación de elevada concentración o incluso de monopolio de un establecimiento comercial perteneciente a otros subsectores.

Un análisis más detallado del subsector comercial de alimentación que según los datos de la *encuesta de comercio interior de 1988* representa, aproximadamente, un 42,5 por ciento del total, nos permite ofrecer algunos datos sobre las cuotas de mercado detentadas por las mayores empresas de alimentación instaladas en nuestro país y la participación de los diferentes formatos comerciales en las ventas totales.

Tomando como cifra de negocios del sector de comercio de alimentación la publicada por el MAPA y considerando que cómo máximo dos tercios de las ventas de las grandes superficies corresponden a productos de alimentación y bebidas[12], las cuotas de las cuatro y de las 10 mayores empresas se muestran en el Cuadro 4.L.3. Su lectura informa de que el grado de concentración en dicho sector es todavía bajo[13], sin embargo está experimentando un claro aumento en los últimos años.

CUADRO 4.L.3. Índices de concentración agregados para el sector de alimentación

Cuotas de mercado	*1990*	*1991*	*1992*	*1993*
C_4	11,9	13,4	15,7	15,8
C_{10}	19,7	22,4	25,6	26,1

Los datos publicados por Nielsen sobre la evolución del comercio de alimentación, referidos, por tanto, a su universo (geográfico, de formas comerciales y de productos) muestran las siguientes evidencias relacionadas con el aumento y la situación actual de la concentración en el sector. En primer lugar, el número de establecimientos comerciales viene decreciendo desde mediados de los setenta, de forma que en 1994 dicho número representa el 76,8 por ciento del correspondiente a 1989; en segundo lugar, el número de establecimientos de gran tamaño (hipermercados) ha experimentado un extraordinario aumento equivalente al 85 por ciento acumulado entre 1989 y 1994; en tercer lugar, la cuota de mercado detentada por los hipermercados casi se ha duplicado entre 1988 y 1993, pasando de representar un 18 a un 30 por ciento de las ventas; en cuarto lugar, las cuatro mayores empresas de hipermercados representan el 89 por ciento de las ventas de este tipo de establecimiento; en quinto lugar, para el año 1993 el 2 por ciento de las mayores empresas representa el 48 por ciento de las ventas y el 10 por el 76 por ciento de las mismas. Por último, la comparación de las cuotas detentadas por los mayores hipermercados españoles frente a los europeos, sitúa a la concentración del sector en nuestro país en las últimas posiciones. Así, mientras la cuota de las tres mayores empresas de alimentación en Finlandia representa el 80 por ciento, en Alemania el 46 por ciento, en Francia y en el Reino Unido el 38 por ciento y en España solo alcanza el 18 por ciento.

4. ANÁLISIS DE LA RELACIÓN ENTRE PRECIOS Y CONCENTRACIÓN EN EL COMERCIO MINORISTA ESPAÑOL

Este análisis empírico tiene por objeto contrastar la relación entre grado de concentración y el poder de monopolio ejercido en el sector de comercio minorista. la información disponible limita el trabajo al estudio del comercio minorista de alimentación seca y a la utilización de medidas *proxies* de las verdaderas variables. Es un trabajo intrasectorial, en el que la dimensión geográfica determina las diferencias de concentración y de resultados analizadas. Es importante tener en cuenta que la información referente al número y tamaño de los establecimientos minoristas de alimentación seca se refiere únicamente a los de mayor tamaño, hipermercados y supermercados de más de 400 metros cuadrados de superficie de venta y no incluye a los autoservicios ni a las tiendas de comercio tradicional.

Datos, metodología y análisis descriptivo

En este estudio se ha utilizado la información referida a los establecimientos de comercio minorista de alimentación seca instalados en 36 capitales de provincia españolas[14]. las fuentes de datos utilizadas son: la revista *Distribución Actualidad* en sus números de 1994 referidos a hipermercados y supermercados, de donde se ha obtenido la información sobre número y superficie de venta de los establecimientos comerciales contemplados en el estudio, así como, información sobre el número de centrales de compra y cadenas de hipermercados y supermercados; la revista *OCU Compra Maestra,* número 164, de la que se utiliza la información sobre índices de precios definidos para dos cestas de compra diferentes, la primera integrada por 113 productos básicos de alimentación e higiene de marcas muy vendidas y con la misma presentación, y la segunda compuesta por 57 productos de primera necesidad o muy corrientes de las marcas más baratas, éstas pueden ser marcas de productor no conocidas o marcas de distribuidor. En esta segunda cesta, por tanto, no coinciden las marcas seleccionadas a lo largo de los establecimientos visitados, por el contrario su nota común es que para cada categoría se elige el precio de la mar-

[12] En el origen de la actividad de los hipermercados, dos tercios correspondían a la venta de productos de alimentación y bebidas. Actualmente hay una tendencia hacia la reducción de esta cuota, de forma que hipótesis creíbles sobre dicha participación situarían la cifra entre el 50 el 70 por ciento. Distribución Actualidad, número 222.

[13] Otras fuentes de información publican cifras de concentración similares. Por ejemplo ALIMARKET situaba la cuota de mercado de los 10 primeros distribuidores de alimentación en España en 1991 en el 18,7 por ciento.

[14] Forman parte de la muestra algunas ciudades que no son capitales de provincia: Vigo, Gijón y Torrelavega, las dos últimas se incluyen como parte de Oviedo y de Santander, respectivamente.

ca que está más barata en el establecimiento. la toma de precios se realizó en las dos primeras semanas del mes de octubre de 1993. Además, se han utilizado otras fuentes estadísticas como la *Renta Nacional de España, 1989,* del BBV, de la que se han tomado las cifras provinciales para 1991 de la renta disponible per cápita; *el anuario del mercado español, 1993,* de Banesto, del que hemos tomado las cifras de población municipal de las capitales de provincia para 1991, y el índice primero de capacidad de compra per cápita; finalmente, hemos utilizado los datos provinciales de la *encuesta de presupuestos familiares de 1991,* referidos al primer grupo de gasto, alimentación, bebidas y tabaco.

Definición y elaboración de variables

La variable dependiente trata de reflejar los resultados, en términos de precios de venta al público, que se alcanzan en los mercados locales de comercio minorista. Para ello tomamos el nivel y la dispersión de precios observados en cada mercado local en la compra de productos de alimentación seca. las medidas concretas son los índices de precios mínimo y máximo recogidos en el estudio de la OCU para cada ciudad y la diferencia entre ellos (PMIN1, PMIN2, PMAX1, PMAX2, DISP1 y DISP2). Aunque para medir el nivel de precios hubiese sido deseable contar con un índice medio, la fuente original no publica información suficiente para su cálculo. Todas las medidas utilizadas tratan de ser *proxies* de la verdadera variable, y por tanto se espera que se aprecie una elevada correlación positiva entre todas ellas. En efecto, en los mercados menos competitivos y con mayor presencia de poder de monopolio se espera que 195 precios de venta al público sean más elevados además que se aprecien mayores diferencias de precios entre las empresas instaladas en el mercado. la definición y elaboración de las variables explicativas se presentan en el Cuadro 4.L.4.

CUADRO 4.L.4. Definición de las variables explicativas

Variable	*Definición*	*Medida*
C_4	Ratio de concentración	Superficie en m^2 de venta de los cuatro mayores-superficie en m^2 de venta del total de hipermercados y supermercados de más de 400 m^2 [15].
TMR	Economías de escala	Tamaño medio relativo medido en superficie de venta. Superficie de venta del establecimiento medio de la ciudad sobre la superficie de venta del establecimiento medio nacional.
CCENTRAL	Poder negociador de los mayoristas	Número de establecimientos abastecidos por la mayor central de compras-número de establecimientos total.
POB	Tamaño de la demanda	Población-m^2 de superficie de venta.
RENR	Capacidad de renta	Renta disponible provincial per cápita-renta disponible nacional per cápita, 1991.
ICC1	Capacidad de compra de productos básicos	Índice primero de capacidad de compra pronvicial per cápita, 1991.
GR	Gasto en productos de alimentación, bebidas y tabaco	Gasto provincial por unidad familiar-gasto nacional por unidad familiar, 1991.

Metodología y análisis descriptivo

En este trabajo, se ha tratado de estimar econométricamente la relación entre algunos indicadores de estructura de mercado, en especial de concentración de mercado, y de resultados, en términos de precios, para el sector de comercio minorista español. Aunque inicialmente se contemplaron seis indicadores de precios alternativos, el análisis econométrico ha forzado la eliminación de los precios mínimos de ambas cestas por no encontrarse ningún modelo que tuviese una suficiente capacidad explicativa[16]. Una posible hipótesis para explicar esta ausencia de relación consiste en el efecto que las promociones pueden haber tenido sobre la toma de datos, que inciden principalmente en los índices mínimos, y que no podemos controlar por el tipo de base de datos utilizada.

Por otra parte, se espera que en los mercados locales en los que los precios máximos son más altos se observe simultáneamente más dispersión de precios. Un sencillo aná-

[15] La indisponibilidad de datos sobre cifra de ventas a nivel de establecimiento, ha forzado a sustituir dicha variable de output por la medida de tamaño más directamente relacionada con ella, la superficie de venta.

[16] Sólo para el caso del precio mínimo de la cesta económica se encuentra alguna variable que explique parcial y limitadamente las diferencias locales observadas en las precios mínimos.

$$PMIN2 = 103{,}74 + 15{,}08 * CCNTRAL$$
$$R^2 = 0{,}21 \quad F = 6{,}40 \quad (32{,}649) \quad (2{,}530)$$

lisis de correlación entre las dos variables de dispersión y sus respectivos índices de precios máximos confirma esta hipótesis[17]. En consecuencia, tanto los índices de precios máximos como la dispersión de precios contienen información del mismo tipo sobre el grado de imperfección de los mercados. Finalmente, cabe esperar que la variable que mide la dispersión de precios de la cesta de marcas conocidas muestre un rango de variación inferior al de la variable que mide la dispersión de precios de la cesta de productos económicos. De un lado, cuando se trata de marcas conocidas existe un mayor control relativo de las condiciones de venta de sus productos, y de otro lado, las marcas más conocidas son utilizadas por los consumidores para comparar, evaluar y seleccionar los establecimientos. En el Cuadro 4.L.5 se presentan algunas estadísticas descriptivas que apoyan estas hipótesis.

CUADRO 4.L.5. Análisis descriptivo

Variable	*Media*	*DT*	*Minimo*	*Máximo*
*PMAX*1	126,19	6,13	115,00	140,00
*PMAX*2	152,29	9,81	137,00	171,00
C_4	0,48	0,16	0,30	0,81
TMR	1,08	0,27	0,69	1,58
CCENTRAL ...	0,49	0,21	0,03	0,84
POB	7,69	3,91	3,14	19,64
RENR	0,98	0,12	0,75	1,23
*ICC*1	0,98	0,09	0,87	1,23
GR	0,98	0,10	0,72	1,21
*DISP*1	21,10	5,92	11,00	34,00
*DISP*2	41,76	11,30	19,00	66,00

Se especifican dos tipos de modelos para cada una de las dos categorías de variables dependientes, *PAMX* y *DISP*. El primero, es lineal y trata de identificar los efectos individuales de los factores explicativos que aproximan el grado de competencia: concentración, economías de escala, y poder negociador de mayoristas y de los factores explicativos de la demanda: tamaño poblacional y capacidad adquisitiva. El segundo, no lineal en las variables de concentración y de economías de escala, trata no sólo de medir los efectos individuales, sino también de averiguar la existencia o no de niveles de concentración y de tamaños óptimos, o la existencia de efectos interacción entre ambos.

Modelo lineal:

$$\text{Precio} = \alpha_0 + \alpha_1 C_4 + \alpha_2 TMR + \alpha_3 POB + \alpha_4 CCOMPRA + \\ + \alpha_5 CCENTRAL + \varepsilon$$

donde Precio puede ser *PMAX*1, *PMAX*2, *DISP*1 o *DISP*2; y donde *CCOMPRA* significa capacidad de compra y puede medirse por *RENR*, *ICC*1 o *GR*.

[17] $DISP1 = 0{,}168 * PMAX1 \quad R^2 = 0{,}95 \quad F = 440{,}5$
$DISP2 = 0{,}276 * PMAX2 \quad R^2 = 0{,}95 \quad F = 446{,}3$

Sobre este modelo las hipótesis a contrastar, tomadas de la clásica relación estructura de mercado–resultados, se concretan en que el signo esperado de todos los coeficientes alfa estimados debe ser positivo.

Modelo no lineal

$$\text{Precio} = \alpha_0 + \alpha_1 C_4 + \alpha_2 C_4^2 + \alpha_3 TMR + \alpha_4 IMR^2 + \\ + \alpha_5 C_4 * TMR + \alpha_6 POB + \alpha_7 CCOMPRA + \\ + \alpha_8 CCENTRAL + \varepsilon$$

Con esta segunda formulación se trata de contrastar las siguientes hipótesis: 1) la concentración Sólo afecta positiva y significativamente al precio en presencia de barreras a la entrada; y 2) el tamaño empresarial y la concentración pueden suponer para los mercados ganancias de eficiencia superiores a las pérdidas potenciales de bienestar derivadas del ejercicio del poder de monopolio hasta un umbral, superado en el cual la relación se invertiría.

La elevada correlación existente entre variables de dimensión y de concentración recomienda utilizar el método de estimación econométrica *stepwise*.

RESULTADOS

Del conjunto de modelos estimados, en sus formas lineales y no lineales, el Cuadro 4.L.6 presenta los seleccionados utilizando criterios de bondad de ajuste estadístico. Su lectura indica que los modelos en los que la variable a explicar es el nivel máximo de precios la varianza explicada es del 40 y del 50 por ciento, respectivamente, según se trate de la cesta de productos de marca o la de productos económicos. En ambos casos, el modelo que mejor se ajusta es lineal en las variables de concentración y de economías de escala, pero mientras las últimas parecen actuar como barreras a la entrada y conducen a precios más altos, la concentración, parece ejercer un efecto negativo sobre el nivel máximo de precios, cualquiera que sea la cesta de productos utilizada para cuantificar la variable precio. Este resultado es coherente con la idea de que en las ciudades donde los grandes hipermercados tienen una participación mayor se introduce una mayor competencia en precios[18]. Por otro lado, en el precio máximo de la cesta de productos de marca la capacidad de compra medida por el gasto relativo ejerce un efecto positivo y significativo. En el precio máximo de productos económicos actúan positiva y significativamente el tamaño de la población y el poder negociador de mayoristas. Mientras que la concentración de las grandes centrales de compra a nivel local no parece ejercer influencia alguna cuando se trata de los precios de productos con marca bien establecida, no ocurre lo mismo cuando la cesta se compone de productos donde los fabricantes delegan al sector de distribución

[18] No hay que olvidar que los datos existentes, publicados por NIELSEN, sobre niveles de precios agregados por formas de establecimiento, ponen de manifiesto que la formulo más económica es el hipermercado.

CUADRO 4.L.6. Resultados de las estimaciones

Variable	*PMAX1*	*PMAX2*	*DISP1*	*DISP2*
C_4	–22,350	–42,988	–19,897	–34,591
	**(–2,633)	***(–3,218)	**(–2,499)	**(–2,105)
TMR	2,056	16,098	44,586	87,116
	**(2,331)	**(2,101)	***(5,700)	***(5,396)
TMR2	–	–	–17,245	–35,102
			***(–4,197)	***(–4,139)
POB	–	0,897	0,508	1,005
		**(2,118)	*(1,937)	*(1,856)
GR	33,365	–	–	
	***(3,055)			
CCENTRAL	–	23,378	–	
		***(3,013)		
CONSTANTE	91,256	137,320	–	–
	***(7,157)	***(14,963)		
R^2	0,410	0,500	0,960	0,956
F	***5,610	***5,950	***128,606	***117,387

Entre paréntesis el estadístico *t*-student.
* Significativo al 90 por ciento; ** 95 por ciento; *** 99 por ciento.

gran parte del control de la comercialización. En este último caso, que en una ciudad los supermercados se abastezcan por la misma o por diferentes centrales de compra tiene un impacto importante, ya que en el primer caso se advierte una reducción importante del nivel de competencia.

Cuando la variable a explicar es la dispersión de precios, el modelo parte de la hipótesis de que en mercados perfectamente competitivos la dispersión observada debe ser nula, por tanto para la estimación adoptamos un ajuste forzado a que la ordenada en el origen sea nula. la varianza explicada en este caso se sitúa en torno al 95 por ciento para ambas cestas y el modelo recoge un efecto no lineal en el tamaño. La concentración sigue mostrando un efecto negativo, que indica que en las ciudades con más concentración se observan menores diferencias entre los índices de precios de los establecimientos más caros y los correspondientes a los más baratos. El efecto del tamaño relativo indica que para la cesta primera (segunda) mientras no se supere el umbral del 1,29 (1,24) la mayor dimensión lleva asociada mayor dispersión, pero que a partir de dichos valores se produce la relación en sentido inverso. la fórmula utilizada para la construcción de esta variable explica en parte este resultado. Los valores más elevados se obtienen cuando existe fuerte concentración o la mayoría de los establecimientos instalados en la ciudad es de grandes dimensiones, por lo que se reducen las diferencias intertipo y se hace más fuerte la competencia entre supermercados e hipermercados, lo que conduce a una menor dispersión de precios.

Finalmente, en ambos modelos se refleja un efecto positivo, de escasa significación estadística, del tamaño de la demanda sobre la dispersión de precios. De forma que en las ciudades donde hay una menor dotación comercial, o más población por metro cuadrado de superficie de venta se observa una mayor dispersión de precios.

REFERENCIAS BIBLIOGRÁFICAS

[1] ARROW, K. (1951): *Social Choice and Individual Values,* Nueva York.

[2] BAIN, J. (1951): «Re1ation of Profit to Industry Concentratio: American Manufacturing, 1936-40», *Quartely journal of Economics.*

[3] BAIN, J. (1956): *Barriers to New Competition,* Harvard University Press.

[4] BMN, J. (1959): *Industrial Organization,* Wiley, Nueva York.

[5] BETANCOURT, R y GAUTSCHI, D. (1988): «The Economics of Retail Firms», *Managerial and Decision Economics,* 19.

[6] DEBREU, G. (1959): *Theor of Value,* Nueva York.

[7] CARREE, M. y THURIK, A. (1990):«Market Volatility, Entry Barriers and Retail Profit Margins», *Report Econometric Institute,* Erasmus University, Rotterdam.

[8] DEIKE, V. (1992): *Análisis de la Variabilidad de los Precios Minoristas en el Sector de la Distribución Comercial,* Comunidad de Madrid, Consejeria de Economía, Madrid.

[9] KOOPMANS, T. (1957): *Tree Essays on the State of Economic Science,* Nueva York.

[10] MULHERN, F. y LEONE, R (1991): «Implicit Priç ce Bundling of Retail Products: A Multiproduct Approach to Maximizing Store Profitability», *Journal of Marketing,* 55.

[11] NELSON, P. e HILKE, J. (1991): «Retail Featuring as a Strategies Entry or Mobility Barrier in Manufacturing», *International Journal of Industrial Organization,* 9.

[12] NOOThBOOM, B. (1985): «A Mark-Up Model of Retail Margins», *Applied Economics,* 17.

[13] PERRY, M. y BESANCO, D. (1991): «Resale Price Maintenance and Manufacturer Competition for Exclusive Dealerships», *The journal of Industrial Economics,* volumen 39, número 5.

[14] SCHERER, F. M. (1980): *Industrial Market Structure and Economic Performance,* Rand Mcnally, Chicago.

[15] SCHERER, F. M. y ROSS, R. (1990): *Industrial Market Structure and Economic Performance,* Houghton Mifflin, Boston.

[16] SCHMALENSEE, R. y WILLIG, R. (1989): *Handbook of Industrial Organization,* North Holland, Amsterdam.

[17] SHAFFER, G. (1991): «Capturing Strategic Rent: Full-Line Forcing, Brand Discount, Aggregate Rebates and Maximun Resale Price Maintenance», *The Journal of Industrial Economics,* volumen 39, número 5.

[18] STEINER, R (1993): «The Inverse Association Between the Margins of Manufacturers and Retailers», *Review of Industrialí Organization,* 8.

[19] VIVES, X. y GUAL, J. (1990): *Concentración Empresarial y Competitividad: España en la CEE,* Ariel Economía.

LECTURA COMPLEMENTARIA

El comportamiento de búsqueda de precios de los consumidores: Un análisis en el mercado minorista de alimentación

CARMEN BERNÉ MANERO, MARTA PEDRAJA IGLESIAS Y PILAR RIVERA TORRES
Universidad de Zaragoza

Revista española de investigación de marketing-ESIC, n.º 1

La revisión de la literatura sobre búsqueda de precios, anima a las autoras del trabajo a realizar un intento de aplicación del modelo desarrollado por Urbany *et al.* (1996) al mercado minorista de alimentación de Zaragoza capital. La aplicación de la metodología de regresión por los métodos de estimación ULS y GLS, proporciona unos resultados muy cercanos a los obtenidos en el estudio que sirve como principal referencia. Entre otras conclusiones, sobresale el grado de conocimiento adquirido sobre precios por el consumidor como el principal factor explicativo del comportamiento de búsqueda de precios.

Palabras clave: *búsqueda de precios; capital humano; beneficios económicos; costes de búsqueda; beneficios psicológicos; características demográficas.*

1. INTRODUCCIÓN

El precio es una de las señales más importantes del mercado ya que está presente en todas las situaciones de compra y representa, para todos los consumidores, el coste monetario mínimo que debe soportar en una transacción.

La teoría económica y la teoría de la percepción son las dos perspectivas de obligada referencia en el análisis de precios. Desde la teoría económica, el precio actúa como mecanismo de asignación de recursos, reflejando la cantidad de dinero a la que un consumidor está dispuesto a renunciar para efectuar una compra, es decir, actúa como una medida de coste o sacrificio. En este sentido, la información más importante que proporciona el precio es el número de unidades de producto que un consumidor podría adquirir antes de agotar plenamente su presupuesto. Desde la perspectiva de la teoría de la percepción, el papel del precio es más complejo, ya que debe ser consistente con el valor percibido del producto. De esta forma, si el consumidor se enfrenta a precios distintos de los que pagaba anteriormente, debe decidir si estas diferencias son o no significativas para él y actuar en consecuencia.

Inmerso en el proceso de compra, y considerando la variable precios, una de las cuestiones a las que se enfrenta el consumidor, es la comercialización de un mismo producto a diferentes precios dependiendo del punto de venta[1] de dicho producto, del momento y de las condiciones de pago. Las diferencias temporales son, tal vez, las que producen una mayor preocupación a corto plazo y es en los productos de alimentación particularmente grave ya que a los factores de estacionalidad de la demanda y a las prácticas comerciales de fabricantes y minoristas se les añade la estaciona-

[1] Considerando el punto de venta, las dispersiones en precios pueden ser intertipo e intratipo, intercanal e intracanal, y dispersiones de precios inter e intracadena. Las diferencias en precios inter e intratipo consideran como criterio de clasificación el nivel de servicio ofertado. Para el primer caso, esto es, dispersiones intertipo, la razón estriba en la existencia de distintos tipos de establecimientos (supermercados, hipermercados...). Las diferencias intratipo se refieren al hecho de que dos establecimientos del mismo tipo nunca son totalmente iguales: simplemente la localización del establecimiento es por sí un servicio. Las dispersiones inter e intracanal se refieren a las diferencias en precios debidas a las distintas formas de vinculación de los establecimientos con sus fuentes de aprovisionamiento, las cuales quedan reflejadas en sus costes de compra y en la forma en que éstos se repercutan a los consumidores. Por último, la dispersión inter e intracadena se debe a las diferencias existentes en la política comercial de las empresas así como a los distintos niveles de control sobre los precios que muestren los establecimientos individuales (Múgica, 1993).

lidad de la oferta (Múgica, 1993). En este sentido, uno de los aspectos sujetos a estudio es si las variaciones de precios a corto plazo (ofertas y promociones), provocan el abandono de la mayor parte de los consumidores de cualquier intento de optimizar su compra de productos de consumo frecuente, por ejemplo, de alimentación. Es decir, si la existencia de incertidumbre en precios en el mercado impide que el consumidor afronte la búsqueda de información necesaria cuando se trata de procesos de compra frecuente, sin caer en costes desproporcionados; tal y como observan Múgica y Deike (1991).

Desde un punto de vista general, la información que necesitan los consumidores para poder evaluar las alternativas existentes en el mercado puede obtenerse a dos niveles: búsqueda interna y nivel búsqueda externa (Beales *et al.*, 1981). El primer nivel supone el esfuerzo de búsqueda inicial que debe realizar el individuo, a menudo de forma inconsciente, acudiendo a sus experiencias pasadas y a sus estructuras de aprendizaje. Si el resultado de esta búsqueda no es suficiente para poder tomar una decisión, deberá pasar al segundo nivel acudiendo al uso de fuentes de información externas. Esta decisión de búsqueda de información externa va a depender de la percepción que tenga el individuo del beneficio que va a obtener frente a los costes (monetarios y no monetarios) que tendrá que asumir si se enfrenta a la búsqueda.

Los trabajos empíricos realizados sobre precios, han utilizado una gran variedad de indicadores de la búsqueda de información externa. Entre ellos se pueden destacar el número de establecimientos visitados, generalmente tres o menos (Newman y Lockeman, 1975; Claxton, Fry y Portis, 1974; Midgley, 1983), el número de marcas evaluadas (Dommermuth, 1965; Furse, Punj y Stewart, 1984), el esfuerzo y tiempo dedicado a la actividad de búsqueda (Kleimenhagen, 1966-67; Newman y Staelin, 1971; Kiel y Layton, 1981; McLelland y Turner, 1983), los recursos de información utilizados (Newman y Staelin, 1972), entre los que se encuentran los anuncios (Bucklin, 1965; Udelí, 1966; Thorelli, 1971; Kiel y Layton, 1981) y la comunicación interpersonal (Udelí, 1966; Thorelli, 1971; Kiel y Layton, 1981; Múgica, 1993), las creencias que posee un individuo sobre la actuación de un determinado mercado (Duncan y Olshavsky, 1982), etc.

En este sentido, hay que considerar una serie de factores económicos que afectan a esta búsqueda de información. Se ha demostrado que el esfuerzo dedicado a la búsqueda de información sobre precios, aumenta con la importancia que tenga el coste de compra del producto dentro del presupuesto familiar (Stigler, 1961), con más bajos límites de aceptabilidad del precio (Múgica, 1993), con el nivel de educación del consumidor (Michael, 1972), con mayores diferencias percibidas en calidades de las marcas (Múgica, 1993), y con la dispersión de precios que exista en el mercado (Stigler, 1961). Por contra, esta búsqueda disminuirá a medida que aumente el nivel de ingresos del consumidor (Linder, 1970; Múgica, 1993), y ante aumentos de los costes temporales ligados a la búsqueda (Mincer, 1963).

Siguiendo a Urbany, Dickson y Key (1991), los consumidores pueden tener buenas razones para variar su propensión a buscar precios en distintos establecimientos minoristas, sin embargo, existe cierta dificultad desde el campo del marketing para comprender y explicar las características más elementales del comportamiento de búsqueda del consumidor en mercados en los cuales la búsqueda de precios constituye una actividad importante: generalmente, se ha sobreestimado la proporción de consumidores que realizan búsquedas activas de precios y que responden a las promociones. Esta observación puede deberse a la inexistencia de estudios que expliquen adecuadamente la búsqueda de precios, por ejemplo, para categorías de productos de compra frecuente, ya que se tiende a examinar series pequeñas de predictores y a utilizar medidas dependientes que no recogen directamente la búsqueda de información sobre precios (Lichtenstein, Ridway y Netemeyer, 1993).

Partiendo de la definición de búsqueda de precios por parte de los consumidores, y de la formulación de unas premisas básicas (Urbany *et al.*, 1996), plantean y contrastan un modelo sobre los determinantes de la búsqueda de precios por los consumidores en un mercado minorista de alimentación, que intenta superar limitaciones de los estudios realizados con anterioridad.

El concepto de búsqueda de precios por parte de los consumidores se define como el esfuerzo empleado en obtener y comparar los precios de los establecimientos competidores (Urbany *et al.*, 1996), obviamente con la esperanza de encontrar, entre otras cosas, precios menores para productos idénticos. La búsqueda de precios se ha identificado con la comparación de precios por los consumidores entre establecimientos y con la búsqueda activa de información relacionada con la compra de los productos y sus precios.

Aceptando la definición, el objetivo de este trabajo consiste fundamentalmente en explicar el comportamiento relativo a la búsqueda de precios del consumidor en productos de alimentación, superando a estudios anteriores al utilizar un número amplio de predictores y mediciones directas de la variable dependiente. Más concretamente, se pretende comprobar la validez predictiva de variables formuladas como independientes sobre la búsqueda de precios de los individuos; así como aplicar las premisas del modelo de los anteriores autores a un mercado minorista de alimentación.

2. DESARROLLO DE LA INVESTIGACIÓN

2.1. Hipótesis del modelo

La investigación desarrollada parte de la formulación de unas hipótesis básicas que relacionan la búsqueda de precios con cinco tipos de factores potencialmente determinantes de la misma. Estos factores son, concretamente, costes de búsqueda, beneficios económicos, capital Humano, características demográficas, y beneficios psicológicos. A continuación se detallan las diferentes hipótesis formuladas[2].

[2] Nótese que, en algunos casos, cada hipótesis general va acompañada de subhipótesis de obligada reformulación por cambios en la dirección propuesta respecto a la variable dependiente.

2.1.1. *Costes de búsqueda*

La actividad de búsqueda requiere incurrir en una serie de costes o hacer frente a restricciones en tiempo disponible, en el proceso de toma de decisiones según las percepciones personales respecto a la complejidad de la tarea de búsqueda, así como posibles restricciones físicas del consumidor. La hipótesis general y las subhipótesis que se plantean son las siguientes:

H1: ***A mayores costes de búsqueda, menor actividad de búsqueda de precios realizará el consumidor***

H1.1: ***A menor tiempo disponible, menor búsqueda de precios***

H1.2: ***A mayor dificultad percibida en la comparación, menor búsqueda de precios***

H1.3: ***A mayores restricciones de movilidad, menor búsqueda de precios***

2.1.2. *Beneficios económicos*

Los beneficios económicos derivados de la actividad de búsqueda se identifican con la consecución del producto a un menor precio. Los consumidores buscarán precios en función de sus percepciones en la distribución del precio (Urbany, 1986; Bucklin, 1969) y en función de la importancia económica de la compra para el consumidor (Stigler, 1961).

H2: ***A mayores beneficios económicos percibidos en la compra, mayor búsqueda de precios***

H2.1: ***A mayor dispersión percibida entre establecimientos, mayor búsqueda de precios***

H2.2: ***A mayores restricciones presupuestarias percibidas, mayor búsqueda de precios***

2.1.3. *Capital humano*

La dimensión de capital humano se refiere fundamentalmente al conocimiento que el individuo acumula con el tiempo. Los consumidores con un conocimiento previo adquirido sobre los precios y su distribución, tienden a limitar la búsqueda en períodos futuros, bajo el supuesto de que los precios están correlacionados en el tiempo (Stigler, 1961).

En este trabajo, se contempla el capital humano tal y como lo hacen Urbany *et al.* (1996), es decir, como el grado de conocimiento actual de los establecimientos competidores, la inversión realizada en la búsqueda en períodos anteriores, que debe permitir simplificar los patrones de compra, por ejemplo identificando el establecimiento preferido y la capacidad para distribuir el tiempo del individuo.

H3.1: ***A mayor grado de conocimiento de los establecimientos locales, menor búsqueda de precios***

H3.2: ***A mayor inversión previa realizada en búsqueda, menor es la posibilidad de búsqueda en períodos futuros***

H3.3: ***A mayor capacidad para distribuir el tiempo, mayor esfuerzo destinado a búsqueda de precios***

2.1.4. *Características demográficas*

Las características demográficas como la edad, el sexo o el nivel educativo han sido utilizadas con frecuencia en las investigaciones como sustitutos de la eficiencia de la búsqueda. Es de esperar que un mayor nivel educativo o una mayor edad proporcionen a los individuos una mayor experiencia como consumidores y, por tanto, sean más eficientes y capaces en su actividad de compra, lo que, en principio, supondría una menor necesidad de búsqueda de precios (Marvel, 1976, Sharir, 1974).

El estudio desarrollado por Urbany *et al.* (1996) no encuentra, sin embargo, que esta dimensión sea un claro determinante de la búsqueda de precios, sobre todo en lo que respecta al sexo.

H4: ***A mayor nivel educativo o mayor edad, menor búsqueda de precios***

2.1.5. *Beneficios psicológicos*

Los beneficios psicológicos en este contexto, tienen que ver por un lado, con el interés de algunos consumidores por ser protagonistas en la recogida de información del mercado y compartirla con los demás, en un intento por demostrar su valía como expertos en el tema en su círculo de amistades o conocidos, algo que se ha venido a denominar «market maven»[3].

Este comportamiento influirá en la búsqueda a través, sobre todo, de su importancia como mecanismo de transmisión de información en el mercado minorista de alimentación (Urbany *et al.*, 1996).

H5.1: ***Un comportamiento de «market maven» estará relacionado positivamente con la búsqueda de precios***

Por otro lado, los individuos consumidores pueden obtener también beneficios psicológicos si entienden que el acto de compra es fundamentalmente un acto de entretenimiento. Esta percepción refleja una utilidad incremental asociada sobre todo a la posibilidad de encontrar precios bajos.

H5.2: ***Una mayor percepción de entretenimiento en la compra estará positivamente asociada con búsqueda de precios***

3. METODOLOGÍA

3.1. Procedimiento utilizado para la obtención de datos

Para desarrollar la investigación, se creó un equipo de trabajo bajo la coordinación de las autoras del mismo, en el que

[3] El «market maven» es un concepto relativo a «aquellos individuos que tienen información sobre muchas clases de productos, lugares donde comprar y otras facetas del mercado, y que comparten dicha información con otros consumidores» (Feick y Price, 1987).

participaron estudiantes de último curso de un instituto de Formación de Profesionales en Técnicos de Marketing y de la Escuela Superior de Gestión Comercial y Marketing (ESIC). Se decidió la utilización de la técnica de encuesta para la recogida de información, elaborándose un cuestionario, que tomaba como base el utilizado por Urbany *et al.* (1996), en el que se recogían las diferentes cuestiones que debían permitir la obtención de los indicadores determinantes de la búsqueda de precios.

La encuesta es personal y el encuestado es el principal responsable de la compra de productos de alimentación en la familia. El mercado geográfico considerado para el estudio es el de Zaragoza capital, ciudad española frecuentemente utilizada como mercado de prueba. Se realizó un muestreo por afijación proporcional teniendo en cuenta la diferente población por distritos de la plaza considerada. Después de la realización de un cuestionario piloto, principalmente con el fin de probar el enunciado de las cuestiones, finalmente el número de encuestas recogidas fue de 231, habiendo resultado válidos 196 cuestionarios, con lo que el error permitido es del 7 por ciento[4].

3.2. Características de la muestra

La muestra está compuesta principalmente por mujeres (el 85,3 por ciento). Al estimar la inexistencia de diferenciación por sexo, no se ha incluido esta variable entre los determinantes de la búsqueda de precios en este estudio[5].

En lo que respecta al resto de características sociodemográficas, el 50 por ciento de la muestra tiene una edad entre los 36 y los 50 años; el 60,5 por ciento perciben unos ingresos brutos anuales de más de un millón y medio, y menos de cuatro millones (el 24,9 entre 1,5 y 2,5 millones y el 35,6 por ciento entre 2,5 y 4 millones respectivamente). La mayor parte de los encuestados (el 90,4 por ciento) poseen estudios, habiendo un reparto equitativo entre aquellos que tenían estudios primarios, secundarios y superiores (31 por ciento, 28 por ciento, y 31 por ciento respectivamente).

3.3. Medición

Prácticamente la totalidad de las cuestiones se planteaban con una escala de puntuación desde 1, completamente en desacuerdo, hasta 5, completamente de acuerdo, a excepción de cuestiones relativas a características demográficas, otros datos personales y frecuencia de compra.

La medición de las variables implicadas en el modelo se especifica en el Cuadro 4.L.1. La mayoría de las mismas son una traducción directa de las utilizadas por Urbany *et al.* (1996), otras son el resultado de la adaptación necesaria observada como resultado del cuestionario piloto, y algunas variables se incluyeron tomando como base el trabajo realizado por Múgica (1993).

[4] La tabla con las necesidades de afijación distrito está a disposición del interesado.

[5] Este hecho se comprobó a través de un test de diferencias de medias.

Como puede observarse en el Cuadro 4.L.1, la medición de la variable dependiente se planteó en dos posibilidades. Una primera, COMPARE, como tendencia a la comparación de precios por el consumidor y, una segunda, ÍNDEX, como búsqueda activa de información sobre precios.

3.4. Procedimiento de análisis

Con el fin de cubrir los objetivos de análisis fijados en el estudio, el procedimiento seguido consistió en la realización de los siguientes pasos: un análisis de fiabilidad, que se abordó a través del cálculo de los coeficientes *Alpha* y las correlaciones *item-to-total;* un análisis factorial exploratorio (utilizando como método de extracción el de *Principal Axis Factoring),* que permitiría observar las diferentes estructuras factoriales de los ítems; y finalmente se aplicaron los métodos de estimación de regresión múltiple ULS y GLS[6], que sirvieron al objetivo de comprobar las hipótesis formuladas, así como la validez predictiva de las variables independientes del modelo.

3.5. Resultados del análisis de fiabilidad y del factorial exploratorio

A efectos de comprobar fiabilidad, se calcularon los coeficientes *alpha* para las dos mediciones de la variable dependiente, búsqueda de precios, obteniéndose unos valores de 0,82 para COMPARE y de 0,74 para ÍNDEX. Igualmente, se calcularon los *coeficientes alpha* para las variables independientes con más de dos ítems, y los coeficientes de correlación para las variables que contaban sólo con dos ítems, obteniéndose los valores que aparecen reflejados en el Cuadro 4.L.2[7].

[6] El método ULS *(unweighted least squares),* es también el utilizado por Urbany *et al.* (1996). Con el objetivo de obtener test estadísticos correctos para los parámetros estimados, dada la no normalidad de las variables, se procedió a la utilización del método de estimación GLS *(generalized least squares),* y se calculó la matriz de covarianzas «robusta» entre los parámetros estimados (Bentler, 1996). Las tablas asociadas a la demostración de la no normalidad de los ítems están a disposición de los interesados.

[7] En las variables incluidas en beneficios psicológicos se eliminaron dos ítems. Concretamente, se observó la conveniencia de eliminar la pregunta 46: 'me gusta probar nuevas marcas y productos», componente de la variable independiente Market Maven, y la pregunta 50: «me planteo el hecho de ir a comprar de una manera positiva», componente de entretenimiento en comprar. Esta decisión producía una ligera mejora de sus alphas de 0,04 y de 0,06. Aunque, en un primer momento, no parecía relevante la eliminación de los ítems, debido fundamentalmente a la intención de no perder información, finalmente se optó por su exclusión ya que, la realización de un análisis factorial exploratorio confirmó un aumento de la varianza explicada con un factor, desde el 64 por ciento hasta el 73, 3 por ciento (en el caso de eliminar el ítem 46), y un aumento de la varianza retenida desde el 70,1 por ciento hasta el 82 por ciento (en el caso de eliminar el ítem 50, cuya estructura era unidimensional). Además, la comunalidad de estas variables era de 0,11 y de 0,28 respectivamente, mientras que el resto de variables, en sus respectivos factores, presentaban unas comunalidades superiores a 0,6 y 0,7 respectivamente.

CUADRO 4.L.1. Variables independientes

Variables independientes

Indicador	*Subindicador*	*N.º Ítem*	*Descripción ítem*
B^{OS} econ.	Dispersión precios (BEDP)	X12 X14 X18 X24	• Un pedido completo de los mismos productos, comprado en distintos establecimientos, cuesta aproximadamente lo mismo • Algunas tiendas en Zaragoza tienen precios mucho más bajos que otras. • El precio de un producto puede variar mucho entre establecimientos. • En X, el precio de los productos cárnicos y agrícolas, varía mucho de un establecimiento a otro.
	Restricción presupuesto (BERP)	X15 X17 X20	• A menudo tengo que gastar más dinero del que pensaba. • Frecuentemente tengo problemas para llegar a fin de mes. • Mi presupuesto es siempre bajo.
	Ingresos U.F. (BEIU)	X61	• Ingresos brutos anuales de la familia.
Ctes. Búsqueda	Restricción tiempo (CBRT)	X16 X28 X59 X60	• Siempre ando falto de tiempo. • Todo lo hago de manera precipitada. • Horas de trabajo remuneradas a la semana. • N.º de niños menores de 5 años.
	Dificultad Comparación (CBDC)	X19 X22	• Me resulta muy difícil comparar la calidad de los productos alimenticios entre establecimientos. • Me resulta muy difícil comparar los precios de diferentes establecimientos.
	Restricción Movilidad (CBRM)	X21 X23 X26 X30	• Mi estado de salud limita mis actividades. • Tengo mucha energía para hacer las cosas. • Puedo disponer de medios de transporte sin problemas. • Tener que utilizar medios de transporte no me condiciona al hacer la compra.
Capital humano	Conocimiento mercado (CHCM)	X25 X27 X29 X31 X32 X33	• Sé mucho sobre los establecimientos de mi ciudad. • Sé que establecimientos tienen los mejores precios. • Sé que establecimientos tienen las mejores ofertas. • Sé que establecimientos tienen la mejor carnicería. • Sé que establecimientos tienen la mejor frutería. • Sé que establecimientos tienen la mejor pescadería.
	Inversión búsqueda previa (CHIB)	X41 X43 X45	• Comparé los precios de diferentes establecimientos antes de decidir dónde realizo la compra ahora. • Realicé un esfuerzo extra al principio para conocer las características de los establecimientos y simplificar mi tarea de compra. • Visité previamente varios establecimientos para identificar aquel donde hago mi compra.
	Cualificación percibida distribución tiempo (CHDT)	X40 X44	• Soy muy hábil en distribuir mi tiempo. • Tengo gran habilidad para organizar y programar mis actividades.
Caract.. Demográf.	Edad (CDED) Educación (CDED)	X57 X58	• Edad. • Educacion.

CUADRO 4.L.1. Variables independientes y dependientes *(continuación)*

Indicador	*Subindicador*	*N.º Ítem*	*Descripción ítem*
Beneficios psicológicos	Market Mavem (BPMM)	X34	• Me gusta que me pregunten sobre dónde conseguir la mejor compra de varios productos.
		X36	• Sé mucho sobre productos, establecimientos y ofertas y me gusta compartir esta información.
		X38	• Pienso en mí mismo como una buena fuente de información para los demás sobre productos y ofertas.
		X46	• Me gusta probar nuevas marcas y productos.
		X47	• Me gusta ayudar a la gente proporcionándole información sobre muchos tipos de productos.
		X48	• Me gusta informar a la gente de los precios de los productos.
		X49	• Me gusta cuando la gente me pregunta sobre productos, lugares donde comprar u ofertas.
	Entretenido comprar (BPEC)	X35	• Hacer la compra es entretenido.
		X37	• Disfruto al hacer la compra.
		X39	• Hacer la compra es divertido.
		X42	• Hacer la compra es un placer.
		X50	• Me planteo el hecho de ir de compras de una forma positiva.

Variables dependientes

Indicador	*N.º Ítem*	*Descripción ítem*
Compare	Xl	• Comparo los precios que ofrecen distintos establecimientos.
	X3	• Normalmente comparo los precios de frutas y vegetales en distintos establecimientos.
	X4	• Normalmente comparo los precios de la carne de dos o más establecimientos.
	X5	• Normalmente comparo los precios del pescado de dos o más establecimientos.
	X6	• Normalmente comparo los precios de los congelados de dos o más establecimientos.
	X2	• Frecuencia con la que compara los precios de diferentes establecimientos.
Índex	X7	• Tengo en cuenta la información que se proporciona en el establecimiento sobre precios de productos sustitutivos o sobre ofertas..., antes de decidir la compra.
	X8	• Generalmente comento con mis amigos las ofertas antes de efectuar la compra.
	X9	• Decido el establecimiento donde voy a comprar en función de la propaganda recibida en casa.
	X10	• Regularmente me informo del establecimiento dónde está el producto en oferta para comprar allí.
	X11	• Selecciono un número de establecimientos a visitar antes de tomar mi decisión de compra.
	X13	• Leo regularmente la propaganda que recibo en mi casa para comparar precios y analizar ofertas.

CUADRO 4.L.2. Coeficientes Alpha. Variables Independientes

Variable	*Alpha/Corre.*
Capital humano: conocimiento mercado (CHCM)	0,84
Capital humano: inversión búsqueda (CHIB)	0,8
Capital humano: distribución tiempo (CHDT)	0,75
Beneficios psicológicos: Market Maven (BPMM)	0,91
Beneficios psicol :entretenimiento comprar (BPEC)	0,87

Los valores de estos coeficientes, o de las correlaciones en su caso, para las variables integrantes de los beneficios económicos (dispersión de precios, restricciones presupuestarias e ingresos por unidad familiar) y de los costes de búsqueda (restricciones temporales, dificultad de comparación y restricciones de movilidad) no se muestran en este cuadro por no haber alcanzado valores aceptables.

Para determinar si realmente existía una estructura común entre los ítems que formaban las variables, tanto dependientes como independientes, se realizó un análisis factorial exploratorio. Es de destacar que para la variable independiente «capital humano: conocimiento del mercado», unidimensional en el planteamiento original, emerge una estructura de dos factores[8] que retienen el 80 por ciento de la varianza. El primer factor está formado por los ítems que indican un «conocimiento general de los precios de los establecimientos», y el segundo factor por aquellos que miden el «conocimiento de las secciones especiales de los establecimientos (pescadería, frutería y carnicería)».

En base a los resultados mencionados, se planteó el modelo que pretendía probar las hipótesis formuladas anteriormente, al igual que lo hicieron Urbany *et al.* (1996). Primeramente, y a diferencia de como se esperaba, no se comprobó una estructura común en los ítems de beneficios económicos y costes de búsqueda (véase Cuadro 4.L.1). Este resultado puede parecer sorprendente, sin embargo, no es lejano al obtenido por Urbany *et al.* (1996)[9]. Aunque estos autores plantean estos dos indicadores como su modelo base, demostrando la existencia de una estructura interna a través del análisis de fiabilidad, lo hacen a razón de su mayor utilización en la literatura anterior. De hecho, el mayor poder explicativo de búsqueda de precios en su modelo, lo tiene el indicador capital humano. Todas estas indicaciones pueden estar reflejando que los ítems utilizados para medir los beneficios económicos y los costes de búsqueda, son más bien indicadores formativos en lugar de reflejos (Bollen y Lennox, 1991).

En cualquier caso, a razón de la observación efectuada, la solución a efectos del estudio actual consistió en no utilizar en la validación posterior del modelo los subindicadores enunciados en el Cuadro 4.L.1, sino directamente los ítems.

Por otra parte, el primer subindicador del capital humano utilizado, esto es, «el grado de conocimiento del mercado», se descompone en dos indicadores (CHCM1 y CHCM2), el primero del «grado de conocimiento del mercado en general» y el segundo «de las secciones especializadas».

3.6. Resultados del análisis de regresión para la variable dependiente COMPARE

En el Cuadro 4.L.3, se presentan los resultados de las regresiones ULS que explican la variable COMPARE. Únicamente se indican los coeficientes beta estandarizados significativos al 95 por ciento.

En la columna A del cuadro, aparecen los resultados obtenidos tomando cada uno de los indicadores de forma independiente. Se observa que tres indicadores -el capital humano, los beneficios psicológicos y los costes de búsqueda- retienen una proporción significativa de la varianza explicada y que es el capital humano el indicador dominante (con un 37 por ciento).

En la columna B, aparecen los resultados de lo que se va a denominar el *modelo base* en este estudio, formado únicamente por el capital humano. Se pretende determinar la aportación del resto de indicadores desde un punto de vista predictivo, lo cual se obtendrá a través de la correlación cuadrática semiparcial[10]. Este modelo base retiene más de un tercio de la varianza en COMPARE (R^2 de 0,389), con tres predictores estadísticamente significativos.

En las columnas C, D, E y F se presentan los resultados obtenidos al añadir al modelo base el resto de indicadores. Ninguno de los resultados obtenidos implican un incremento importante de la explicación de la variable dependiente COMPARE. La mayor aportación aparece en la columna E, a excepción del modelo completo (Columna G), donde la incorporación de los ítems representativos de los costes de búsqueda, implican un aumento del R^2 de 0,046 respecto al obtenido en el modelo base. Los predictores significativos son «la inversión en búsqueda previa realizada», «el estado de salud» y «el tener que utilizar medios de transporte para la realización de la compra».

Por otra parte, también se debe destacar que ni lo beneficios psicológicos ni las características demográficas son estadísticamente significativos en ninguno de los modelos planteados, por lo que no contribuyen a aumentar la explicación de la tendencia de los consumidores a comparar precios.

En la columna G del Cuadro 4.L.3, se presentan los resultados obtenidos para el modelo completo, obteniéndose un R^2 de 0,435, siendo los predictores estadísticamente significativos «el estado de salud», «tener mucha energía para hacer las cosas» y «la inversión realizada en búsqueda previa».

Finalmente, en la última columna del cuadro, se presentan los resultados correspondientes a la eliminación, dentro del modelo completo, del indicador «capital humano». De esta forma, se observa que el poder explicativo de la variable COMPARE que tiene el capital humano, es superior al que presentan el resto de indicadores del modelo cuando se utilizan conjuntamente.

---

[8] Este resultado coincide con el obtenido por Urbany et al (1996).

[9] Hay que indicar que ya Goldman y Johansson, en 1978, se plantearon la utilidad de la teoría de la economía de la información para explicar el comportamiento de compra de gasolina. Estos autores, sin realizar ningún análisis de fiabilidad, llegaron a resultados muy débiles en nivel de significatividad, lo que les llevó a cuestionar la función básica de esta teoría: *búsqueda = f(ingresos marginales, costes marginales)*.

[10] La correlación cuadrática semiparcial de un indicador cuando se analizan dos modelos de regresión, se calcula simplemente restando del coeficiente de determinación del modelo completo la varianza explicada por el modelo que incluye únicamente el otro indicador: $sR^2_A = R^{2Y.AB} - R^2_{YB}$ (Cohen y Cohen, 1975).

CUADRO 4.L.3. Resultados de las regresiones para la variable dependiente COMPARE

Variable	*A Serie indiv.* Beta AdjR²	*B Modelo base*	*C Base+ Beneficios Psicológ.*	*D Base + ítems beneficios económicos*	*E Base + ítems costes búsqueda*	*F Base + Caract. Demogr.*	*G Modelo completo*	*H Modelo completo capital humano*
* *Beneficios económicos:*	**0,08**							
X12: Pedido completo cuesta igual en distintas tiendas								
X14: Muchas tiendas tienen precios más bajos que otras								
X11: El precio de un producto varía mucho entre establecimientos								
X24: El precio de los prdtos agrícolas y cárnicos varía mucho entre tiendas	0,285			0,161				0,189
X15: A menudo gasto más de lo que pensaba				(0,117)				
X17: Frecuentemente tengo problemas para llegar a fin de mes								
X20: Mi presupuesto es siempre bajo								
X61: Ingresos brutos anuales unidad familiar								
* *Costes de búsqueda:*	**0,109**							
X16: Siempre ando falto de tiempo								
X28: Todo lo hago de forma precipitada								
X59: Horas de trabajo remuneradas a la semana								
X60: N.º de niños menores de 5 años								
X19: Dificultad para comparar calidad de los productos								
X22: Dificultad comparar precios entre establecimientos								
X21: Salud limita las actividades	0,268				0,208		0,886	
X23: Energía para hacer las cosas	0,199						0,61	
X26: Disponer de transporte sin problemas								
X30: No condiciona el uso de transporte para comprar	0,261				0,143			0,205
* *Capital humano:*	**0,37**							
CHCM1: Conocimiento mercado en general	0,183	0,183				0,182		
CHCM2: Conecimiento mercado por secciones especiales								
CHIB: Inversión búsqueda previa	0,411	0,411	0,398	0,401	0,405	0,404	0,692	
CHDT: Cualificación percibida distribución del tiempo	0,147	0,147	0,128			0,145		
* *Características demográficas:*	**0,03**							
CDED: Edad	0,218							
CDEU: Educación								
**Beneficios psicológicos:*	**0,149**							
BPMM: Market Maven	0,414							0,29
BPEC: Entretenimiento comprar								
R^2 ajustado del modelo		**0,376**	**0,379**	**0,409**	**0,423**	**0,572**	**0,435**	**0,240**
Variación del R^2 ajustado sobre el modelo base			0,004	0,032	0,046	−0,004	0,056	−0,063

Por tanto, a modo de resumen, se puede concluir afirmando que la tendencia que muestran los consumidores a comparar precios viene determinada por el capital principalmente, por «el conocimiento de los establecimientos en general», por «la inversión previa realizada en la búsqueda» y por «la capacidad del individuo para distribuir su tiempo». Estos tres predictores muestran una relación significativa y positiva con la variable dependiente COMPARE.

3.7. Resultados del análisis de regresión para la variable dependiente ÍNDEX

En el Cuadro 4.L.4, se presentan los resultados de las regresiones ULS que explican la variable ÍNDEX. Hay que indicar que, igual que en el caso anterior, únicamente se indican los coeficientes beta significativos al 95 por ciento.

En la columna A del cuadro, aparecen los resultados obtenidos tomando cada uno de los indicadores de forma independiente. Se observa que tres indicadores —el capital humano, los beneficios psicológicos y los beneficios económicos— retienen una proporción significativa de la varianza explicada (29,2 por ciento, 23 por ciento, y 14 por ciento respectivamente). La proporción de varianza explicada por los indicadores es superior que en el caso anterior (veáse Cuadro 4.L.3), excepto para el capital humano. Este hecho indica que el poder predictivo de estos indicadores es mayor a la hora de determinar la búsqueda activa de información sobre precios que para analizar la tendencia a la comparación de precios. También hay que destacar que las características demográficas no son significativas en la búsqueda activa de precios.

En la columna B, aparecen los resultados de lo que se va a denominar el *modelo base,* formado únicamente por el capital humano. En este caso, el coeficiente de determinación ajustado es de 0,299 presentando dos predictores estadísticamente significativos: «el conocimiento general de los precios de los establecimientos» y «la inversión realizada en búsqueda previa».

En las columnas C, D, E y F del Cuadro 4.L.4, se presentan los resultados obtenidos al añadir al modelo base el resto de indicadores. Se observa que ninguno de ellos aumenta considerablemente el poder explicativo que tiene el capital humano sobre la búsqueda activa de precios. Además, se mantiene la pauta, observada ya en el modelo base (columna B): son «el conocimiento general de los precios de los establecimientos» y «la inversión en búsqueda previa», los predictores positivos y estadísticamente significativos de la variable dependiente ÍNDEX.

En la columna G se presentan los resultados del modelo completo. La inclusión de todos los indicadores proporciona un incremento del poder explicativo respecto al conseguido por el modelo base: el coeficiente de determinación ajustado es ahora de 0,392. En este caso, los coeficientes estadísticamente significativos son los correspondientes a las variables independientes: «conocimiento general de los precios de los establecimientos» y comportamiento «market maven» de los consumidores, presentando la primera variable un coeficiente mayor que la segunda.

Por último, la columna H muestra los resultados obtenidos al excluir del modelo completo el capital humano, que presenta un R^2 ajustado de 0,306. Este resultado indica que, al contrario de lo que ocurría cuando la variable dependiente era COMPARE, el poder explicativo de la variable ÍNDEX del resto de indicadores es ligeramente superior al que presenta el capital humano. En este caso, las variables con coeficientes estadísticamente significativos son un ítem representativo de «la dispersión de precios» y el comportamiento de «market maven» de los consumidores.

4. REFLEXIONES SOBRE LOS RESULTADOS OBTENIDOS PARA LAS VARIABLES DEPENDIENTES

Son varias las conclusiones que pueden extraerse en relación a las hipótesis planteadas inicialmente. En primer lugar, las hipótesis generales planteadas (Hl e H2: «a mayor coste de búsqueda, menos búsqueda de precios» y «a mas beneficios económicos, más búsqueda de precios», respectivamente), no se pueden comprobar, debido a la inexistencia de una estructura común entre los ítems utilizados en su medición. Por este motivo, únicamente son susceptibles de comprobación las subhipótesis planteadas (H1.1, H1.3; H2.1 y H2.2), ya que existen ítems estadísticamente significativos que reflejan tanto los beneficios económicos conseguidos con la búsqueda de precios en el sector de la alimentación, como los costes de búsqueda en los que incurren los consumidores.

En lo que respecta al resto de hipótesis planteadas, hay que analizar los resultados obtenidos de forma independiente para las variables dependientes COMPARE (tendencia a la comparación de precios) e ÍNDEX (búsqueda activa de precios). A este esfuerzo se dedican las siguientes líneas.

La hipótesis H1.1 se acepta para la variable ÍNDEX, ya que el ítem 59 presenta un coeficiente negativo y estadísticamente significativo. Es decir, se comprueba que a mayor número de horas trabajadas remuneradas a la semana, se realiza una menor búsqueda activa de precios, lo que demuestra la importancia de las restricciones temporales a la hora de que los consumidores busquen activamente precios menores en este mercado.

La hipótesis H1.3 planteaba que «a más restricciones de movilidad, menor búsqueda de precios». En este sentido, aparecen tres ítems estadísticamente significativos: «el estado de salud», «la energía de los consumidores» y «el tener que utilizar medios de transporte», como no condicionantes de la compra.

Para las dos variables dependientes utilizadas, se comprueba que «a más energía para realizar las cosas, más búsqueda de precios» y que «el tener que utilizar medios de transporte», no hace que los consumidores realicen una menor búsqueda de precios, ya que ambas presentan coeficientes positivos y significativos tanto en lo que se refiere a la tendencia a comparar precios como a la búsqueda activa de precios menores. El ítem restante, «mi estado de salud limita mis actividades» presenta un resultado inesperado, ya que su coeficiente estimado es significativo y positivo res-

CUADRO 4.L.4. Resultados de las regresiones para la variable dependiente INDEX

Variable	*A Serie indiv.* Beta AdjR²	*B Modelo base*	*C Base+ Beneficios Psicológ.*	*D Base + ítems beneficios económicos*	*E Base + ítems costes búsqueda*	*F Base + Caract. Demogr.*	*G Modelo completo*	*H Modelo completo capital humano*
* *Beneficios económicos:*	**0,138**							
X12: Pedido completo cuesta igual en distintas tiendas								
X14: Muchas tiendas tienen precios más bajos que otras								
X11: El precio de un producto varía mucho entre establecimientos								
X24: El precio de los prdtos agrícolas y cárnicos varía mucho entre tiendas	0,279			0,199				0,195
X15: A menudo gasto más de lo que pensaba								
X17: Frecuentemente tengo problemas para llegar u fin de mes								
X20: Mi presupuesto es siempre bajo								
X61: Ingresos brutos anuales unidad familiar	(0,193)			(0,2)				
* *Costes de búsqueda:*	**0,082**							
X16: Siempre ando falto de tiempo								
X28: Todo lo hago de forma precipitada								
X59: Horas de trabajo remuneradas a la semana	(0,215)				(0,167)			
X60: N.º de niños menores de 5 años								
X19: Dificultad para comparar calidad de los productos								
X22: Dificultad comparar precios entre establecimientos								
X21: Salud limita las actividades	0,169							
X23: Energía para hacer las cosas	0,296							
X26: Disponer de transporte sin problemas								
X30: No condiciona el uso de transporte para comprar	0,168							
* *Capital humano:*	**0,299**							
CHCM1: Conocimiento mercado en general	0,419	0,419	0,321	0,409	0,398	0,414	0,588	
CHCM2: Conecimiento mercado por secciones especiales								
CHIB: Inversión búsqueda previa	0,211	0,211	0,184	0,163	0,184	0,224		
CHDT: Cualificación percibida distribución del tiempo								
* *Características demográficas:*								
CDED: Edad								
CDEU: Educación								
Beneficios psicológicos:	**0,229**							
BPMM: Market Maven	0,506		0,277				0,173	0,4
BPEC: Entretenimiento comprar								
R^2 ajustado del modelo		**0,299**	**0,339**	**0,361**	**0,321**	**0,307**	**0,392**	**0,306**
Variación del R^2 ajustado sobre el modelo base			0,004	0,062	0,023	0,008	0,093	0,007

pecto a las dos variables dependientes. Esto puede deberse a que el ítem se haya planteado en términos muy generales. En cualquier caso, lo obtenido concluye que cuanto más limitado se considera el individuo en la realización de sus actividades, más búsqueda activa de precios realiza y mayor tendencia a la comparación de precios manifiesta. Las limitaciones de salud pueden estar ligadas a la no realización de otras tareas, como el trabajo, deporte, u otras actividades de ocio ligadas a movilidad física, sin embargo, sí que pueden dar lugar a un mayor interés por el individuo, por conseguir eficiencia en sus decisiones de compra. Además, las dos variables a explicar son variables que en su medición no consideran la realización de esfuerzos físicos elevados.

La hipótesis H2.1 se refiere a la dispersión de precios existente en el mercado. En este sentido, emergió un ítem estadísticamente significativo en las dos variables dependientes. Este ítem se refiere a las secciones de especialidades existentes en los establecimientos, por lo que se puede afirmar que «a mayor dispersión de precios de los productos agrícolas y cárnicos existente en el mercado, mayor búsqueda de precios realizarán los consumidores», lo cual permitiría aceptar la hipótesis.

Por otra parte, se comprueba que los ingresos brutos anuales de la unidad familiar condicionan la tendencia de los consumidores a comparar precios. El coeficiente negativo y estadísticamente significativo obtenido para este ítem conduce a aceptar la hipótesis H2.2. Es decir, «a mayores ingresos brutos anuales obtenidos por la unidad familiar, existe una menor búsqueda activa de precios entre los establecimientos». Este resultado viene a corroborar el obtenido anteriormente en otras investigaciones anteriores realizadas para el mercado español en un contexto similar (véase Múgica, 1993).

En lo que se refiere a los componentes del modelo base (capital humano), planteado para las dos variables dependientes, se obtienen los siguientes resultados: en la hipótesis H3.1, se planteaba que «a mayor conocimiento del mercado, menor búsqueda de precios». Partiendo de que este conocimiento del mercado se divide en dos, un conocimiento general de los establecimientos (CHCM1), y «un conocimiento de las secciones especiales -agrícola, cárnica y pescadería-» (CHCM2), sólo el primero de ellos muestra un coeficiente positivo tanto para la variable COMPARE como para ÍNDEX. Consecuentemente, no se puede aceptar la H3.1, lo que indica que, los consumidores con un mayor conocimiento de los precios de los establecimientos mostrarán en general una mayor tendencia a la comparación de precios y una mayor búsqueda activa de precios. Lo mismo sucede con respecto a la H3.2. En base a los resultados obtenidos, esta hipótesis no puede ser aceptada. De manera que, cuanto mayor haya sido la inversión realizada en búsqueda de precios en períodos anteriores al del análisis, mayor es la búsqueda de precios en la actualidad. La explicación puede estar en que la experiencia de los consumidores en búsqueda, encuentren mayores utilidades en el seguimiento de dicha actividad.

En este orden de cosas, las hipótesis planteadas, aunque son consistentes con la teoría del capital humano establecida por Stigler y Becker (1977), no tienen en cuenta que la información sobre los precios (principalmente en este mercado de productos perecederos), se renueva en cortos períodos de tiempo, lo que hace que los consumidores tengan la necesidad de ir actualizando su información sobre el mercado en general. Por tanto, los resultados obtenidos vienen a demostrar que los consumidores con un mayor conocimiento del mercado son más eficientes y, por tanto, pueden estar realizando búsqueda de precios con unos costes percibidos menores. Esta apreciación se complementa con la aceptación de la hipótesis H3.3, para el caso de la variable COMPARE: en efecto, aquellos consumidores con una mayor habilidad para distribuir su tiempo disponible, realizan una mayor búsqueda de precios.

Si se analizan las características demográficas de los consumidores: nivel de educación y edad, se observa, por un lado, que para educación no se consiguen coeficientes significativos con ninguna de las variables dependientes. En cambio, la edad muestra una relación positiva y estadísticamente significativa con la tendencia de los consumidores a comparar precios. Esto lleva a rechazar la H4 planteada, ya que una mayor edad de los encuestados se correlaciona positivamente con una mayor búsqueda de precios. Este resultado coincide con el obtenido por Urbany *et al.* (1996), quienes a través de la realización de un análisis multivariante, concluyeron que al contrario que los consumidores jóvenes, los más mayores realizan una mayor búsqueda de precios, principalmente por sus menores restricciones de tiempo, su mayor conocimiento del mercado y sus mayores beneficios psicológicos percibidos derivados de la búsqueda. En este mismo sentido, Múgica (1993), comprueba que, bajo la consideración de que los consumidores de mayor edad tienen una mayor preocupación por los precios, éstos realizarán menos actividades de búsqueda laboriosa y más aquéllas que requieran un menor esfuerzo.

Por último, hay que destacar la importancia del carácter de «market maven» que muestren los consumidores a la hora de buscar precios. Se observa que, tanto para la tendencia a la comparación de precios, como para la búsqueda activa de dichos precios, este indicador muestra un coeficiente positivo y estadísticamente significativo, lo que lleva a la aceptación de la H5.1. Además, como ya se ha comentado anteriormente, los beneficios psicológicos, cuyo principal predictor es el «market maven», presentan un R^2 ajustado muy similar al mostrado por el modelo base en el caso de la variable ÍNDEX, lo que demuestra su capacidad explicativa de la búsqueda activa de menores precios en el mercado de productos alimenticios.

5. CONCLUSIONES, LIMITACIONES E INVESTIGACIONES FUTURAS

A modo de resumen, y recordando que el principal objetivo de esta investigación consistía en la comprobación de hipótesis formuladas en relación a búsqueda de precios, se presentan a continuación las principales conclusiones extraídas:

5.1. A diferencia de lo que rezan referencias anteriores en la literatura de marketing especializada en el análisis de

búsqueda de precios, ni los ítems de beneficios ni los de costes económicos percibidos por los consumidores, presentan en esta investigación una estructura común que permita su inclusión como indicadores reflejos, confirmando la necesidad de contemplarlos como indicadores formativos.

5.2. El mayor poder predictivo, tanto para explicar la tendencia a la comparación de precios como para analizar la búsqueda activa de información, lo presenta el capital humano. Este resultado corrobora el obtenido por Urbany *et al.* (1996). Dentro del capital humano, el «grado de conocimiento del mercado en general» y «la inversión previa realizada en la búsqueda de precios», son los principales predictores de la búsqueda de precios. Resultado que corrobora, por un lado, el hecho de que los consumidores con un mayor conocimiento del mercado y con una mayor inversión previa en búsqueda son más eficientes en la realización de esta actividad y, por otro lado, la necesidad que el consumidor observa en los mercados como el que aquí se analiza, de estar al día en la información sobre los precios de los productos.

5.3. Los beneficios psicológicos también presentan un poder predictivo significativo, tanto en la tendencia a comparar precios mostrada por los consumidores como en la búsqueda activa de información sobre precios. Sin embargo, para ambos casos, únicamente presenta un coeficiente positivo y significativo el carácter de «market maven» del consumidor.

5.4. De las características demográficas consideradas en el análisis, es la edad la que presenta un coeficiente significativo y positivo, aunque únicamente cuando la variable a explicar recoge la tendencia a la comparación de precios. Esto es, no se constata la relación cuando la variable a explicar es búsqueda activa de precios. Por tanto, se puede afirmar que una mayor edad del consumidor supone una mayor tendencia a la comparación de precios.

En base a estas conclusiones se pueden obtener implicaciones para decisiones operativas y estratégicas de marketing. Por un lado, parece aconsejable la diferenciación en precios de los productos perecederos, principalmente cárnicos y agrícolas, diferenciación que se podría gestionar a través de promociones continuadas en el tiempo, puesto que se observan efectos a corto plazo en la búsqueda de precios de aquellos consumidores que perciben dispersiones en precios de este tipo de productos.

Por tanto, la utilización por los establecimientos de estas políticas de precios, además de ser una herramienta de marketing ofensivo dirigida a captar clientes, podría derivar en una herramienta de marketing defensivo mediante una aplicación continuada. Hay que tener en cuenta que, por una parte, aquellos consumidores que manifiesten un comportamiento de «market maven» se encargarán de compartir con el resto de consumidores la información que obtengan del mercado y, por otra, que la eficiencia de los consumidores a la hora de buscar precios va aumentando con la experiencia y dichos consumidores seguirán obteniendo información sobre los precios de los distintos establecimientos existentes en el mercado. Por tanto, aunque el efecto de las promociones de precios se mantiene únicamente en un corto período de tiempo, dichas estrategias realizadas de manera continua pueden conseguir la fidelización de este tipo de consumidores.

Por otro lado, también es factible la diferenciación del establecimiento a través del servicio. Con un mayor nivel en la relación servicio-precio, el comerciante se dirigirá a rentas altas, consumidores jóvenes y con menor tiempo disponible. Con este perfil de consumidor, la tendencia a la comparación de precios es menor y, por tanto, más adecuadas las políticas de marketing defensivo encaminadas, desde las primeras decisiones de gestión comercial, a la fidelización del cliente.

Para finalizar, se desea destacar que el trabajo aquí presentado pretende ser el inicio de una línea de investigación a desarrollar sobre el comportamiento de búsqueda de precios de los consumidores.

Más concretamente, es de interés de las autoras la ampliación del estudio expuesto en varios sentidos. Por un lado, se desea profundizar en las relaciones entre las variables consideradas independientes, por ejemplo, mediante la utilización del análisis discriminante, esperando obtener una mayor y mejor información sobre aspectos que se han observado en este primer estudio. Por otro lado, se ha de ser consciente de que el estudio se ha realizado sin diferenciar los tipos de dispersión en precios. En este sentido, se cuenta con información sobre diferencias inter e intratipo que serán susceptibles de estudio en trabajos posteriores. Así mismo, otro tipo de variables son potenciales entrantes como variables explicativas del comportamiento de búsqueda, como la búsqueda de variedad intrínseca del individuo o la separación de los diferentes costes de búsqueda (de percepción, prospección y evaluación), que tienen que soportar los consumidores.

BIBLIOGRAFÍA

BEALES, H., MAZIS, B., SALOP; S. C. y STAELIN, R. (1981). «Consumer Search and Public Policy», *Journal of Consumer Research.* Vol. 8.

BENTLER, P. M. (1996). «Covariance Structure Analysis: Statistical Practice, Theory and Directions», *Anual Review Psychological.* Vol. 47.

BUCKLIN, L. P. (1965). «The Informative Role of Advertising», *Journal of Advertising Research.* Vol. 5, n.º 3.

BUCKLIN, L. P. (1969). «Consumer Search, Role Enactment and Market Efficiency», *Journal of Business,* Vol. 42.

CLAXTON, J., FRY, N. y PORTIS, B. (1974). «A Taxonomy of Purchase Information Gathering Patterns», *Journal of Consumer Research.* Vol. 1.

COHEN, J., COHEN, P. (1975). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* New York. John Wiley & Sons.

DOMMERMUTH, W. P. (1965). «The Shopping Matrix and Marketing Strategy», *Journal of Marketing Research.* Vol. 2, Mayo.

DUNCAN, **C.** P. y OLSHAVSKY, R. W. (1982). «External Search: The Role of Consumer Beliefs», *Journal of Marketing Research.* Vol. 19, Febrero.

FEICK, L. F. y PRICE, L. L. (1987). «The Market Maven: A Diffuser of Marketplace Information», *Journal of Marketing.* Vol. 51, Enero.

FURSE, D. H., PUNJ, G. N. y STEWART, D. W. (1984). «A Typology of Individual Search Strategies Among Purchasers of New Automobiles», *Journal of Consumer Research.* Vol. 19, nº 4.

GOLDMAN, A., JOHANSSON, J. K (1978). «Determinants of Search for Lower Prices: An Empirical Assessment of the Economics of Information Theory», *Journal of Consumer Research.* Vol. 5, Diciembre.

KLEIMENHAGEN, A. K. (1966-67). «Shopping Speciality or Convenience Goods», *Journal of Retailing.* Vol. 42, Invierno.

KIEL, G. C. y LAYTON, R. A. (1981). «Dimensions of Consumer Information Seeking Behavior», *Journal of Marketing Research.* Vol. 18, Mayo.

LCHTENSTEIN, D. R., RIDWAY, N. M. y NETEMEYER, R. G. (1993). «Price Perceptions and Consumer Shopping Behavior», *Journal of Marketing Research.* Vol. 30, May.

LINDER, 5. B. (1970). *The Harried Leisure Class.* Edit. Columbia University Press, New York.

MARVEL, H. P. (1976). «The Economics of Information and Retail Gasoline Price Behavior: An Empirical Analysis», *Journal of Political Economy.* Vol. 84.

MCLELLAND, C. y TURNER, J. (1983). «Consumer Prepurchase Information Seeking: A Comparison of Methodologies», *Marketing Intelligence and Planning.* Vol. 1.

MICHAEL, R. T. (1972). *The Effect of Education on Efficiency in Consumption.* New York: National Bureau of Economic Search.

MIDGLEY, D. F. (1983). «Patterns of Interpersonal Information Seeking for the Purchase of a Simbolic Product», *Journal of Marketing Research.* Vol. 20, Febrero.

MINCER, J. (1963). *Market Prices, Opportunity Cost and Income Effects; in Measurement en Economies. Studies in Mathematical Economics and Econometrics.* California: Standford University Press.

MÚGICA GRIJALBA, J. M. (1993). «El Consumidor ante los Precios de los Bienes No Duraderos: Evaluación del Esfuerzo de Búsqueda e Identificación de los Precios Internos de Referencia»,

Memoria de Cátedra: proyecto de investigación. No publicado.

MÚGICA, J. M. y DEIKE, V. (1991). «La Variabilidad de Precios: Problemas en la Generación y Difusión de Información a los Consumidores», *Esic. Actas del III Encuentro de Profesores Universitarios de Marketing,* Salamanca.

NEWMAN, J W y LOCKEMAN, B. D. (1975). «Measuring Prepurchase Information Seeking», *Journal of Consumer Research.* Vol. 2.

NEWMAN, J W y STAELIN, R. (1971). «Multivariate Analysis of Differences in Buyer Decision Time», *Joarnal of Consumer Research.* Vol. 8, Mayo.

NEWMAN, J W y STAELIN, R. (1972). «Prepurchase Information Seeking for New Car and Major Household Appliances», *Journal of Marketing Research,* Vol. 9, Agosto.

SHARIR, 5. (1974). «Brand Loyalty and the Houdehold's Cost of Time», *Joarnal of Business.* Vol. 57.

STIGLER, G. (1961). «The Economics of Information», *Journal of Political Economy* nº 69, Junio.

STIGLER, G., BECKER, G. S. (1977). 'De Gustibus Non Est Dispuntamdum», *American Economic Review,* Vol. 67, Marzo.

THORELLI, H. B. (1971). «Concentration of Information Power Among Consumers», *Journal of Marketing Research,* Vol. 8, Noviembre.

UDELL, J. G. (1966). «Prepurchase Behavior of Buyers of Small Electrical Appliances», *Journal of Marketing,* Vol. 30, Octubre.

URBANY, J. E. (1986). «An Experimental Examination of the Economics of Information», *Journal of Consumer Research,* Vol. 13, Septiembre.

URBANY, J. E., DICKSON, P. R. y KEY, R. (1991). «Actual and Perceived Consumer Vigilance in the Retail Grocery Industry», *Marketing Letters,* Vol. 2, nº 1.

URBANY, J. E., DICKSON, P. R. y KALAPURAKAL, R. (1996). «Price Search in the Retail Grocery Market», *Journal of Marketing,* Vol. 60, Abril.

CAPÍTULO 5

Análisis discriminante múltiple y regresión logística

OBJETIVOS DE APRENDIZAJE

- Establecer bajo qué circunstancias es mejor utilizar una función discriminante lineal en lugar de la regresión múltiple.
- Comprender los supuestos básicos del análisis discriminante para valorar la conveniencia de su uso en un problema particular.
- Identificar las cuestiones principales en la aplicación del análisis discriminante.
- Describir los dos métodos de cálculo para el análisis discriminante y establecer cuándo debe emplearse cada uno.
- Cómo interpretar la naturaleza de la función discriminante lineal, es decir, identificar las variables independientes con capacidad discriminante significativa.
- Explicar la utilidad de la metodología de la matriz de clasificación y cómo construir una matriz de clasificación.
- Describir los diferentes métodos para evaluar el poder clasificatorio de la función discriminante, incluyendo la distinción entre el ratio de aciertos y el R^2 de la regresión múltiple.
- Justificar el uso de un enfoque de división de la muestra para validar la función discriminante.
- Comprender los puntos fuertes y débiles del análisis discriminante en comparación con la regresión logística.
- Interpretar los resultados de un análisis de regresión logística, en comparación con la regresión múltiple y el análisis discriminante.

Sin duda, la regresión múltiple es la técnica de dependencia multivariante utilizada más extensamente. El principal factor de la popularidad de la regresión ha sido su capacidad para predecir y explicar las variables métricas. Pero, ¿qué decir sobre las variables no métricas? La regresión mul-

tivariante no es adecuada en este contexto. Este capítulo presenta dos técnicas —el análisis discriminante y la regresión logística— que tratan la situación cuando la variable dependiente es no métrica. En esta situación, el investigador está interesado en la predicción y explicación de las relaciones que influyen en la categoría en que un objeto está situado, como por ejemplo, por qué una persona es o no cliente, o si una empresa tendrá éxito o fracasará. Este capítulo evita la jerga y las fórmulas matemáticas sin ocultar importantes conceptos. El capítulo tiene dos objetivos principales: (1) introducir la naturaleza subyacente, la filosofía y las condiciones del análisis discriminante múltiple y la regresión logística; y (2) demostrar la aplicación e interpretación de estas técnicas con un ejemplo ilustrativo.

En el Capítulo 1 se estableció que el propósito básico del análisis discriminante es estimar la relación entre una única variable dependiente no métrica (categórica) y un conjunto de variables independientes métricas, en esta forma general:

$$\underset{\text{(no métrica)}}{Y_1} = \underset{\text{(métricas)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

El análisis discriminante múltiple y la regresión logística cuentan con amplias aplicaciones en situaciones donde el primer objetivo es identificar el grupo al cual un objeto (por ejemplo, una persona, una empresa, o un producto) pertenece. Las aplicaciones potenciales incluyen la predicción de éxitos o fracasos de un nuevo producto, decidir si un estudiante debe ser admitido en una universidad, clasificar a los estudiantes por sus intereses vocacionales, determinar en qué categoría de riesgo de crédito se encuentra una persona o predecir si una empresa tendrá éxito. En cada caso los objetos están incluidos en grupos y se desea que la pertenencia a cada grupo de cada objeto pueda predecirse o explicarse por un conjunto de variables independientes seleccionadas por el investigador con una de las dos técnicas presentadas en este capítulo.

¿Qué son el análisis discriminante y la regresión logística?

Al intentar elegir una técnica analítica apropiada, algunas veces encontramos un problema que incluye una variable dependiente categórica y varias variables independientes métricas. Por ejemplo, podemos querer distinguir entre riesgo de crédito alto y bajo. Si tuviéramos una medida métrica del riesgo de crédito, podríamos utilizar la regresión multivariante. Pero puede ocurrir que sólo podamos conocer si alguien se encuentra en una categoría de riesgo bueno o malo. Esta no es la medida de tipo métrico requerida para el análisis de regresión múltiple.

El análisis discriminante y la regresión logística son las técnicas estadísticas apropiadas cuando la variable dependiente es categórica (nominal o no métrica) y las variables independientes son métricas. En muchos casos, la variable dependiente consta de dos grupos o clasificaciones, por ejemplo, masculino frente a femenino o alto frente a bajo. En otras situaciones, se incluyen más de dos casos, como en una clasificación de tres grupos que comprenda clasificaciones bajas, medias y altas. El análisis discriminante tiene la capacidad de tratar tanto dos grupos como grupos múltiples (tres o más). Cuando se incluyen dos clasificaciones, la técnica es conocida como análisis discriminante de dos grupos. Cuando se identifican tres o más clasificaciones, la técnica es conocida como análisis discriminante múltiple (*MDA*). La **regresión logística**, también conocida como **análisis logit**, está restringida en su forma básica a dos grupos, aunque en formulaciones alternativas puede considerar más de dos grupos.

El análisis discriminante implica obtener un **valor teórico**, es decir, una combinación lineal de dos (o más) variables independientes que discrimine mejor entre los grupos definidos a priori. La discriminación se lleva a cabo estableciendo las ponderaciones del valor teórico para cada varia-

ble de tal forma que maximicen la varianza entre-grupos frente a la varianza intra-grupos. La combinación lineal para el análisis discriminante, también conocida como **función discriminante**, se deriva de una ecuación que adopta la siguiente forma:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + W_3X_{3k} + \cdots + W_nX_{nk}$$

donde

Z_{jk} = puntuación Z discriminante de la función discriminante j para el objeto k
a = constante
W_i = ponderación discriminante para la variable independiente i
X_{ik} = variable independiente i para el objeto k

El análisis discriminante es la técnica estadística apropiada para contrastar la hipótesis de que las medias de los grupos de un conjunto de variables independientes para dos o más grupos son iguales. Para realizar esto, el análisis discriminante multiplica cada variable independiente por su correspondiente ponderación y suma estos productos. El resultado es una única **puntuación *Z* discriminante** compuesta para cada individuo en el análisis. Promediando las puntuaciones discriminantes para todos los individuos dentro de un grupo particular, obtenemos la media del grupo. Esta media del grupo es conocida como **centroide**. Cuando el análisis engloba dos grupos, existen dos centroides; con tres grupos, hay tres centroides, y así sucesivamente. Los centroides indican la situación más común de cualquier individuo de un determinado grupo, y una comparación de los centroides de los grupos muestra lo apartados que se encuentran los grupos a lo largo de la dimensión que se está contrastando.

El contraste para la significación estadística de la función discriminante es una medida generalizada de la distancia entre los centroides de los grupos. Se calcula comparando las distribuciones de las puntuaciones discriminantes para los grupos. Si el solapamiento en la distribución es pequeño, la función discriminante separa bien los grupos. Si el solapamiento es grande, la función es un mal discriminador entre los grupos. Las distribuciones de las puntuaciones discriminantes mostradas en la Figura 5.1 ilustran mejor este concepto. El diagrama de arriba representa la distribución de puntuaciones discriminantes para una función que separa bien los grupos, mientras que el diagrama de abajo muestra la distribución de puntuaciones discriminantes de una función que es un discriminador relativamente malo entre los grupos A y B. Las áreas sombreadas representan la probabilidad de clasificar erróneamente objetos del grupo A en el B.

El análisis discriminante múltiple resulta único en una característica sobre la relación de dependencia que aquí nos interesa: si existen más de dos grupos en la variable dependiente, el análisis discriminante calculará más de una función discriminante. En realidad, calculará $NG - 1$ funciones, donde NG es el número de grupos.

Cada función discriminante calculará una puntuación discriminante Z. En el caso de una variable dependiente de tres grupos, cada objeto tendrá una puntuación para las funciones discriminantes una y dos, permitiendo que los objetos puedan dibujarse en dos dimensiones, donde cada dimensión representa una función discriminante. De esta forma, el análisis discriminante no está limitado a un sólo valor teórico, como la regresión múltiple, sino que obtiene valores teóricos múltiples que representan dimensiones de discriminación entre los grupos.

La regresión logística es un tipo especial de regresión que se utiliza para predecir y explicar una variable categórica binaria (dos grupos) en lugar de una medida dependiente métrica. La forma del valor teórico de la regresión logísitica es similar a la del valor teórico en la regresión múltiple. El valor teórico representa una única relación multivariante con coeficientes como los de la regresión que indican la influencia relativa de la variable predictor. Las diferencias entre la regresión logística y el análisis discriminante resultarán más claras cuando presentemos las características específicas de la regresión logística más adelante en este capítulo. Aún así existen también muchas semejanzas entre ambos métodos. Cuando se conocen los supuestos básicos de ambas, estas técnicas proporcionan resultados predictivos y clasificatorios comparables y emplean medidas de validación similares. Sin embargo la regresión logística tiene la ventaja de verse menos afectada que el análisis discriminante cuando no se cumplen los supuestos básicos, concretamente la

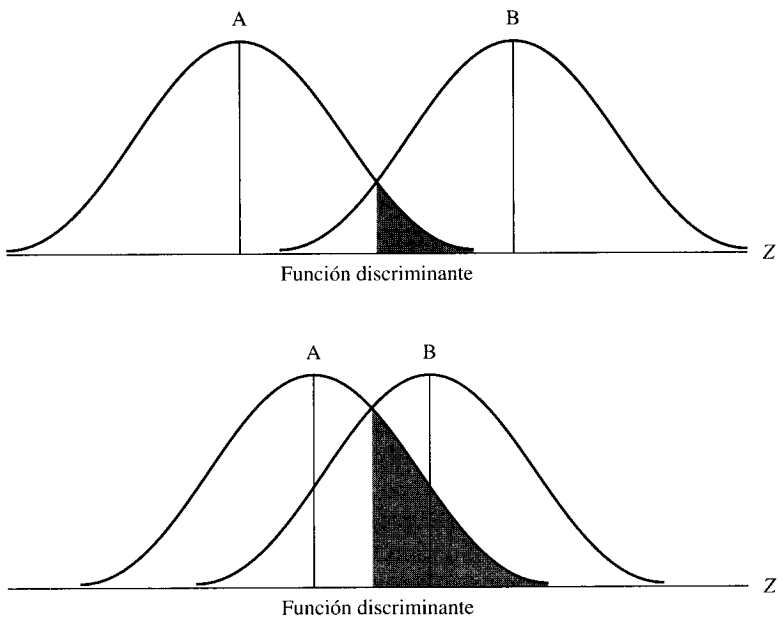


FIGURA 5.1. Representación univariante de las puntuaciones *Z* discriminantes.

normalidad de las variables. Además puede permitir la utilización de variables no métricas por medio de su codificación con variables ficticias, tal como puede hacerse en la regresión. La regresión logística está limitada, sin embargo, a la predicción de tan sólo la medida dependiente de dos grupos. Por tanto, en casos donde la medida dependiente está formada por dos o más grupos se adecúa mejor el análisis discriminante.

Analogía con la regresión y MANOVA

La aplicación e interpretación del análisis discriminante es parecida al análisis de regresión; es decir, la función discriminante es una combinación lineal (valor teórico) de medidas métricas de dos o más variables independientes y se usa para describir o predecir una única variable dependiente. La principal diferencia es que el análisis discriminante es apropiado en trabajos de investigación en donde la variable dependiente es categórica (nominal o no métrica), mientras que la regresión se utiliza cuando la variable dependiente es métrica. Como se señaló antes, la regresión logística es una variante de la regresión, contando con semejanzas excepto por el tipo de variable dependiente. El análisis discriminante también es comparable a «invertir» el análisis multivariante de la varianza (MANOVA), que presentamos en el Capítulo 6. En el análisis discriminante, la variable dependiente única es categórica, y las variables independientes son métricas. El caso opuesto es el de MANOVA, donde se incluyen variables dependientes métricas y variable(s) independiente(s) categórica(s).

Un ejemplo hipotético de análisis discriminante

El análisis discriminante se puede aplicar a cualquier problema de investigación que tenga por objetivo la comprensión de la pertenencia a un grupo, donde los grupos comprenden individuos (por ejemplo, clientes frente a no clientes), empresas (por ejemplo, rentables frente a no rentables), productos (por ejemplo, de éxito en ventas frente a fracaso en ventas), o cualquier otro objeto que pueda evaluarse sobre un conjunto de variables independientes. Para ilustrar los supuestos básicos del

análisis discriminante, examinamos dos ámbitos de investigación, uno que incluye dos grupos (compradores frente a no compradores) y otro de tres grupos (niveles o comportamiento de cambio). La regresión logística opera de forma muy parecida al análisis discriminante de dos grupos. No ilustramos concretamente aquí la regresión logística, aplazando la discusión de la regresión logística para más adelante en este capítulo.

Análisis discriminante de dos grupos: Compradores frente a no compradores

Supongamos que HATCO quiere averiguar si uno de sus nuevos productos —una nueva y mejorada batidora— tendrá éxito comercial. Para llevar a cabo la investigación, HATCO en primer lugar está interesada en identificar (si es posible) a aquellos consumidores que comprarían el nuevo producto y a aquellos que no lo comprarían. En terminología estadística, HATCO querría minimizar el número de errores que se cometerían al predecir qué consumidores comprarían la nueva batidora y cuáles no. Para ayudar a identificar a los compradores potenciales, HATCO ha ideado escalas de valoración para tres características —duración, funcionamiento y diseño— que son utilizadas por los consumidores para evaluar el nuevo producto. En lugar de considerar cada escala como una medida separada, HATCO espera que una combinación ponderada de las tres mejore la predicción de si es probable que un consumidor compre el nuevo producto.

El análisis discriminante puede obtener una combinación ponderada de las tres escalas que se utilice para predecir la verosimilitud de que un consumidor compre el producto. Además de determinar si los clientes que seguramente compren el nuevo producto se pueden distinguir de los que no lo comprarán, HATCO querría saber qué características de su nuevo producto son útiles para diferenciar a los compradores de los no compradores; es decir, ¿cuál de las tres características del nuevo producto separa mejor a los compradores de los no compradores? Por ejemplo, si la respuesta «compraría» esta siempre asociada con una valoración de la duración alta y la respuesta «no compraría» esta siempre asociada con una valoración de la duración baja, HATCO concluirá que la característica de duración diferenciaría a los compradores de los no compradores. Por el contrario, si HATCO descubre que muchas personas con una valoración alta sobre el diseño comprarían la batidora tanto como aquellas que no, entonces el diseño sería una característica que discrimina pobremente entre compradores y no compradores.

La Tabla 5.1 refleja las valoraciones sobre esas tres características de la nueva batidora con un precio especificado por un grupo de 10 compradores potenciales. En la valoración de la batidora, cada miembro de un panel podría estar comparándola implícitamente con productos ya existentes en el mercado. Después de que el producto sea evaluado, se pidió a los evaluadores que establecieran sus intenciones de compra («compraría» o «no compraría»). Cinco determinaron que comprarían la nueva batidora y cinco dijeron que no.

La Tabla 5.1 identifica potencialmente varias variables discriminantes. Lo primero es que existe una diferencia sustancial entre las valoraciones medias de los grupos «compraría» y «no compraría» sobre X_1, la característica de la duración (7,4 – 3,2 = 4,2). La duración parece que discrimina bien entre los grupos «compraría» y «no compraría» y es probable que sea una característica importante para los compradores potenciales. Por otro lado, la característica de diseño (X_3) presenta una diferencia mucho más pequeña de 0,2 entre las valoraciones medias (4,0 – 3,8 = 0,2) para los grupos «compraría» y «no compraría». Por tanto, esperaríamos que esta característica sea menos discriminante en términos de la decisión de comprar o no. Sin embargo, antes de que podamos llegar a conclusiones tajantes, debemos examinar las distribuciones de las puntuaciones para cada grupo. Grandes desviaciones estándar dentro de uno o de ambos grupos pueden hacer que la diferencia entre medias sea no significativa e inoperante para discriminar entre grupos.

Ya que solamente tenemos 10 encuestados en dos grupos y tres variables independientes, también observaremos los datos gráficamente para determinar qué análisis discriminante se intentará llevar a cabo. La Figura 5.2 muestra a los 10 encuestados sobre cada una de las tres variables. El

TABLA 5.1. Resultados de la encuesta HATCO sobre la evaluación de un nuevo producto de consumo

Grupos basados en la intención de compra	*Evaluación del nuevo producto**		
	X_1 *Duración*	X_2 *Funcionamiento*	X_3 *Diseño*
Grupo 1: Compraría			
Sujeto 1	8	9	6
Sujeto 2	6	7	5
Sujeto 3	10	6	3
Sujeto 4	9	4	4
Sujeto 5	4	8	2
Media del grupo	7,4	6,8	4,0
Grupo 2: No compraría			
Sujeto 6	5	4	7
Sujeto 7	3	7	2
Sujeto 8	4	5	5
Sujeto 9	2	4	3
Sujeto 10	2	2	2
Media del grupo	3,2	4,4	3,8
Diferencias entre las medias de los grupos	4,2	2,4	0,2

* Las evaluaciones están hechas sobre una escala de 10 puntos (1 = muy baja a 10 = excelente).

grupo «compraría» está representado por círculos y el grupo «no compraría» por cuadrados. Los números de identificación de los encuestados están dentro de las figuras. Observando primero X_1 (la duración) que cuenta con una una diferencia sustancial en las puntuaciones medias, vemos que podríamos discriminar casi perfectamente entre los grupos usando solamente esta variable. Si estableciéramos el valor de 5,5 como punto de corte para discriminar entre los dos grupos, entonces

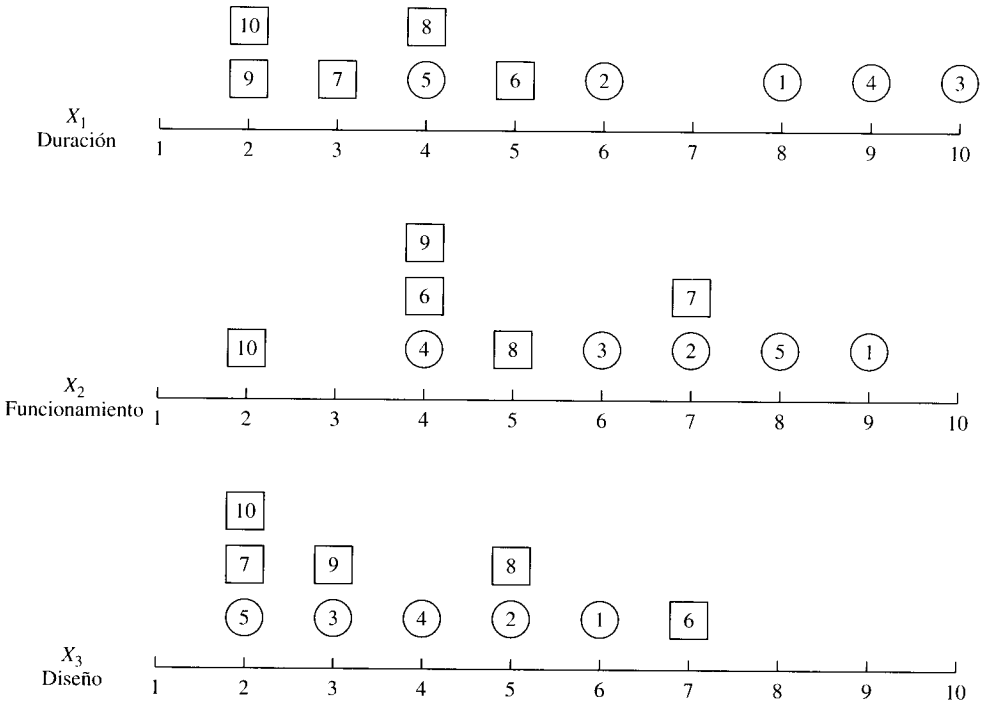


FIGURA 5.2. Representación gráfica de 10 compradores potenciales sobre tres posibles variables discriminantes.

clasificaríamos de forma incorrecta al encuestado número cinco, uno de los miembros del grupo «compraría». Esto es indicativo de la gran diferencia en las medias de los dos grupos en lo que se refiere a duración y de la ausencia de solapamiento entre las distribuciones de los dos grupos (véase Figura 5.1). Examinando X_2 (el funcionamiento), vemos que hay una distinción menos clara entre los dos grupos. Sin embargo, esta variable proporciona una alta discriminación para el encuestado número cinco, que fue mal clasificado al usar X_1. Además, los encuestados que estuviesen mal clasificados usando X_2 están bien separados con X_1. Por ello X_1 y X_2 podrían utilizarse de forma bastante eficiente en combinación para predecir la pertenencia a un grupo. Finalmente el diseño, X_3, muestra pequeñas diferencias entre los grupos. Por ello, construyendo un valor teórico con X_1 y X_2 y omitiendo X_3 se podría formar una función discriminante que maximizara la separación de los grupos en base a la puntuación discriminante.

La Tabla 5.2 contiene los resultados de tres funciones discriminantes diferentes. La primera función discriminante contiene sólo X_1, igualando el valor de X_1 a la puntuación Z discriminante. Como se discutió anteriormente, el uso exclusivo de X_1, el mejor discriminador, implicaba la clasificación errónea del sujeto 5. Como se muestra en la **matriz de clasificación** en la Tabla 5.2, cuatro de los cinco sujetos en el grupo 1 (todos excepto el quinto) y cinco de los cinco sujetos del grupo 2 son clasificados correctamente (esto es, caen en la diagonal de la matriz de clasificación). El **porcentaje correctamente clasificado** es por tanto del 90 por ciento (9 de cada 10 sujetos). Debido a que X_2 proporciona discriminación para el sujeto 5, podemos formar una segunda función discriminante combinando igualmente X_1 y X_2 para utilizar las capacidades discriminatorias únicas de cada variable. Como vemos en la Tabla 5.2, usando una puntuación de corte de 11 con esta nueva función discriminante alcanzamos una clasificación perfecta de los dos grupos (100 por cien correctamente clasificado). Por tanto, la combinación de X_1 y X_2 es capaz de hacer las mejores predicciones de pertenencia al grupo que cada una de ellas separadamente.

El análisis discriminante sigue un procedimiento muy similar al que está indicado en el anterior ejemplo hipotético al estimar empíricamente la función discriminante. Identifica los variables

TABLA 5.2. Construcción de funciones discriminantes para predecir a los compradores y a los no compradores

	Puntuaciones Z discriminantes calculadas		
Grupo	*Función 1:* $Z = X_1$	*Función 2:* $Z = X_1 + X_2$	*Función 3:* $Z = -4{,}53 + 0{,}476X_1 + 0{,}359X_2$
Grupo 1: Compraría			
Sujeto 1	8	17	2,51
Sujeto 2	6	13	0,84
Sujeto 3	10	16	2,38
Sujeto 4	9	13	1,19
Sujeto 5	4	12	0,25
Grupo 2: No compraría			
Sujeto 6	5	9	−0,71
Sujeto 7	3	10	−0,59
Sujeto 8	4	9	−0,83
Sujeto 9	2	6	−2,14
Sujeto 10	2	4	−2,86
Puntuación de corte	5,5	11	0,0

PRECISIÓN AL CLASIFICAR:

	Grupo predicho	
Grupo real	1	2
1: Compraría	4	1
2: No compraría	0	5

Grupo predicho	
1	2
5	0
0	5

Grupo predicho	
1	2
5	0
0	5

clasificaríamos de forma incorrecta al encuestado número cinco, uno de los miembros del grupo «compraría». Esto es indicativo de la gran diferencia en las medias de los dos grupos en lo que se refiere a duración y de la ausencia de solapamiento entre las distribuciones de los dos grupos (véase Figura 5.1). Examinando X_2 (el funcionamiento), vemos que hay una distinción menos clara entre los dos grupos. Sin embargo, esta variable proporciona una alta discriminación para el encuestado número cinco, que fue mal clasificado al usar X_1. Además, los encuestados que estuviesen mal clasificados usando X_2 están bien separados con X_1. Por ello X_1 y X_2 podrían utilizarse de forma bastante eficiente en combinación para predecir la pertenencia a un grupo. Finalmente el diseño, X_3, muestra pequeñas diferencias entre los grupos. Por ello, construyendo un valor teórico con X_1 y X_2 y omitiendo X_3 se podría formar una función discriminante que maximizara la separación de los grupos en base a la puntuación discriminante.

La Tabla 5.2 contiene los resultados de tres funciones discriminantes diferentes. La primera función discriminante contiene sólo X_1, igualando el valor de X_1 a la puntuación Z discriminante. Como se discutió anteriormente, el uso exclusivo de X_1, el mejor discriminador, implicaba la clasificación errónea del sujeto 5. Como se muestra en la **matriz de clasificación** en la Tabla 5.2, cuatro de los cinco sujetos en el grupo 1 (todos excepto el quinto) y cinco de los cinco sujetos del grupo 2 son clasificados correctamente (esto es, caen en la diagonal de la matriz de clasificación). El **porcentaje correctamente clasificado** es por tanto del 90 por ciento (9 de cada 10 sujetos). Debido a que X_2 proporciona discriminación para el sujeto 5, podemos formar una segunda función discriminante combinando igualmente X_1 y X_2 para utilizar las capacidades discriminatorias únicas de cada variable. Como vemos en la Tabla 5.2, usando una puntuación de corte de 11 con esta nueva función discriminante alcanzamos una clasificación perfecta de los dos grupos (100 por cien correctamente clasificado). Por tanto, la combinación de X_1 y X_2 es capaz de hacer las mejores predicciones de pertenencia al grupo que cada una de ellas separadamente.

El análisis discriminante sigue un procedimiento muy similar al que está indicado en el anterior ejemplo hipotético al estimar empíricamente la función discriminante. Identifica los variables

TABLA 5.2. Construcción de funciones discriminantes para predecir a los compradores y a los no compradores

	Puntuaciones Z discriminantes calculadas		
Grupo	*Función 1:* $Z = X_1$	*Función 2:* $Z = X_1 + X_2$	*Función 3:* $Z = -4{,}53 + 0{,}476X_1 + 0{,}359X_2$
Grupo 1: Compraría			
Sujeto 1	8	17	2,51
Sujeto 2	6	13	0,84
Sujeto 3	10	16	2,38
Sujeto 4	9	13	1,19
Sujeto 5	4	12	0,25
Grupo 2: No compraría			
Sujeto 6	5	9	−0,71
Sujeto 7	3	10	−0,59
Sujeto 8	4	9	−0,83
Sujeto 9	2	6	−2,14
Sujeto 10	2	4	−2,86
Puntuación de corte	5,5	11	0,0

PRECISIÓN AL CLASIFICAR:

	Grupo predicho		Grupo predicho		Grupo predicho	
Grupo real	1	2	1	2	1	2
1: Compraría	4	1	5	0	5	0
2: No compraría	0	5	0	5	0	5

con las diferencias más grandes entre los grupos y obtiene un coeficiente de ponderación para cada variable para reflejar estas diferencias. La tercera función discriminante en la Tabla 5.2 representa la verdadera función discriminante estimada ($Z = -4{,}53 + 0{,}476X_1 + 0{,}359X_2$). Empleando una puntuación de corte de 0. La tercera función también alcanza un porcentaje de clasificación correcta del 100 por cien con la máxima separación posible entre grupos.

Una representación geométrica de la función discriminante de dos grupos

Una ilustración gráfica de otro análisis de dos grupos ayudará a explicar mejor la naturaleza del análisis discriminante [6]. La Figura 5.3 demuestra lo que ocurre cuando se calcula una función discriminante de dos grupos. Supongamos que tenemos dos grupos, A y B, y dos medidas, V_1 y V_2, para cada miembro de los dos grupos. Podemos dibujar en un diagrama de dispersión la relación de la variable V_1 con la variable V_2 para cada miembro de los dos grupos. En la Figura 5.3 los puntos pequeños representan las medidas de las variables para los miembros del grupo B y los grandes para los del grupo A. Las elipses alrededor de los puntos grandes y pequeños encerrarían algunas proporciones preespecificadas de los puntos, generalmente el 95 por ciento o más en cada grupo. Si dibujamos una línea recta a través de los dos puntos donde las elipses se cortan y proyectamos

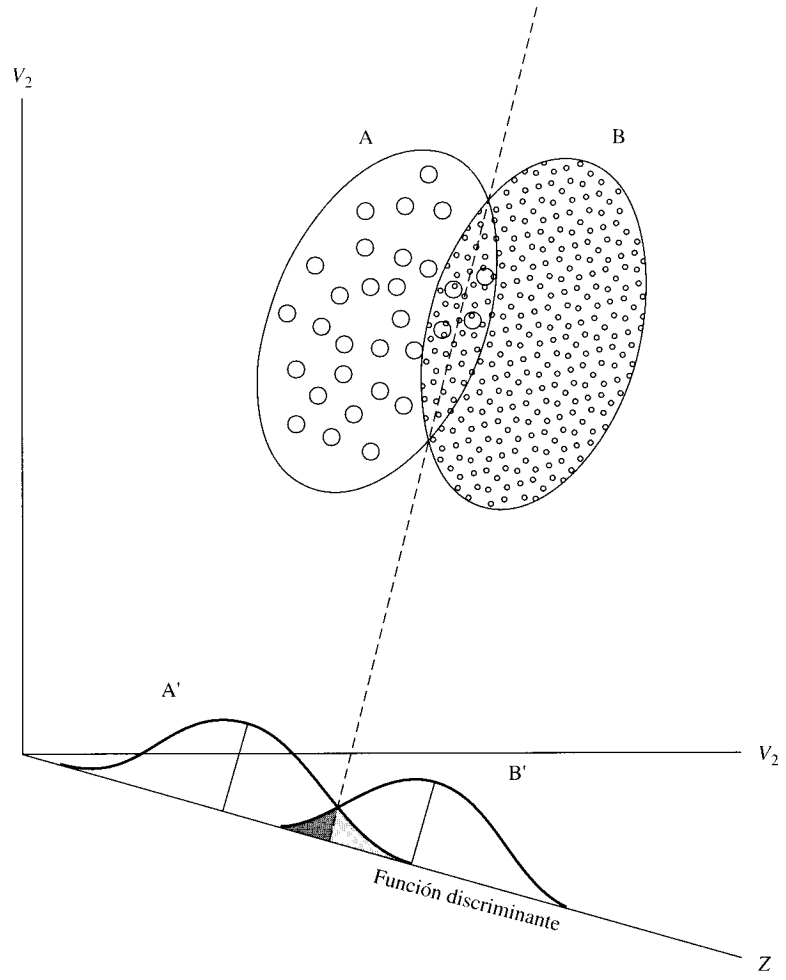


FIGURA 5.3. Ilustración gráfica del análisis discriminante de dos grupos.

la línea a un nuevo eje Z, podemos decir que la superposición entre las distribuciones univariantes A′ y B′ (representada por la superposición de las áreas sombreadas) es más pequeña que la que se obtendría si se dibujase cualquier otra línea a través de las elipses formada por los diagramas de dispersión [6].

Lo que conviene destacar de la Figura 5.3 es que el eje Z representa los perfiles de dos variables de los grupos A y B como únicos números (puntuaciones discriminantes). Encontrando una combinación lineal de las variables originales V_1 y V_2, podemos proyectar los resultados como una función discriminante. Por ejemplo, si los puntos grandes y pequeños se proyectan sobre el nuevo eje Z como puntuaciones Z discriminantes, el resultado resume la información sobre las diferencias de los grupos (mostrado en el plano V_1V_2) dentro de un conjunto de puntos (puntuaciones Z) sobre un único eje, mostrado por las distribuciones A′ y B′.

Para resumir, en un determinado problema de análisis discriminante, se halla una combinación lineal de las variables independientes, obteniéndose una serie de puntuaciones discriminantes para cada objeto en cada grupo. Las puntuaciones discriminantes se calculan de acuerdo a la regla estadística de maximizar la varianza entre los grupos y minimizar la varianza intra grupos. Si la varianza entre los grupos es grande con relación a la varianza intra grupos, diremos que la función discriminante separa bien los grupos.

Un ejemplo de análisis discriminante de tres grupos: Propósitos de cambio

El ejemplo de dos grupos recién examinado demuestra la razón de ser y las ventajas de combinar variables independientes en un valor teórico para discriminar entre grupos. Pero el análisis discriminante también cuenta con otros medios de discriminación: la estimación y el uso de valores teóricos múltiples en casos donde hay tres o más grupos. En estos casos, estas funciones discriminantes llegan a ser dimensiones de discriminación: cada dimensión se separa y se diferencia de las otras. Por ello, además de mejorar la explicación de la pertenencia a un grupo, estas funciones discriminantes adicionales aportan ideas entre las varias combinaciones de variables independientes que discriminan entre los grupos.

Como ilustración de una aplicación del análisis discriminante de tres grupos, examinaremos la investigación llevada a cabo por HATCO referente a la posibilidad de que los clientes de un competidor cambien de proveedor. En una primera encuesta a pequeña escala se entrevistó a 15 clientes de un competidor importante. A lo largo de la entrevista se les preguntó a los clientes por la probabilidad de cambiar de proveedor en una escala de tres categorías. Las tres respuestas posibles fueron «con toda seguridad cambiar», «indeciso» y «con toda seguridad no cambiar». Los clientes fueron asignados a los grupos 1, 2 y 3, respectivamente, de acuerdo con su respuesta. Los clientes también valoraron al competidor en las características de competitividad en el precio y nivel de servicio. El objetivo de la investigación se centra ahora en determinar si las valoraciones de los clientes sobre su proveedor habitual pueden predecir la probabilidad de abandonar a ese proveedor. Dado que la variable dependiente de la probabilidad de cambiar se midió como una variable categórica (no métrica) y las valoraciones del precio y del servicio son métricas, el análisis discriminante es apropiado.

Con tres categorías para la variable dependiente, el análisis discriminante puede estimar dos funciones discriminantes, cada una representando una dimensión diferente de discriminación. La Tabla 5.3 contiene los resultados del estudio para los 15 clientes, 5 en cada categoría de la variable dependiente. Al igual que hicimos en el ejemplo de dos grupos, podemos examinar las puntuaciones medias de cada grupo para ver si una de las variables discrimina bien entre todos los grupos. Para X_1, la competitividad en el precio, vemos una diferencia en la media bastante más grande entre el grupo 1 y los grupos 2 o 3 (0,0 frente a 1,6 y 2,25). X_1 puede discriminar bien entre el grupo 1 y los grupos 2 o 3 pero probablemente será mucho menos efectiva para discriminar entre los grupos 2 y 3. Para X_2, el nivel de servicio, vemos que la diferencia entre los grupos 1 y 2 es muy pequeña (2,0 frente a 2,2), mientras que existe una gran diferencia entre el grupo 3 y los grupos 1

TABLA 5.3. Resultados de la encuesta de HATCO sobre las intenciones de cambio de compradores potenciales

	*Evaluación del proveedor actual**	
Grupos basados en la intención de cambio	X_1 *Competitividad en el precio*	X_2 *Nivel de servicio*
Grupo 1: Se cambiarán con seguridad		
Sujeto 1	2	2
Sujeto 2	1	2
Sujeto 3	3	2
Sujeto 4	2	1
Sujeto 5	2	3
Media del grupo	2,0	2,0
Grupo 2: Indecisos		
Sujeto 6	4	2
Sujeto 7	4	3
Sujeto 8	5	1
Sujeto 9	5	2
Sujeto 10	5	3
Media del grupo	4,6	2,2
Grupo 3: No cambiarán con seguridad		
Sujeto 11	2	6
Sujeto 12	3	6
Sujeto 13	4	6
Sujeto 14	5	6
Sujeto 15	5	7
Media del grupo	3,8	6,2

* Las evaluaciones están hechas sobre una escala de 10 puntos (1 = muy baja a 10 = excelente).

y 2 (6,2 frente a 2,0 y 2,2). Así, X_1 diferencia el grupo 1 de los grupos 2 y 3, mientras que X_2 diferencia el grupo 3 de los grupos 1 y 2. Como resultado, vemos que X_1 y X_2 proporcionan diferentes «dimensiones» de discriminación entre los grupos.

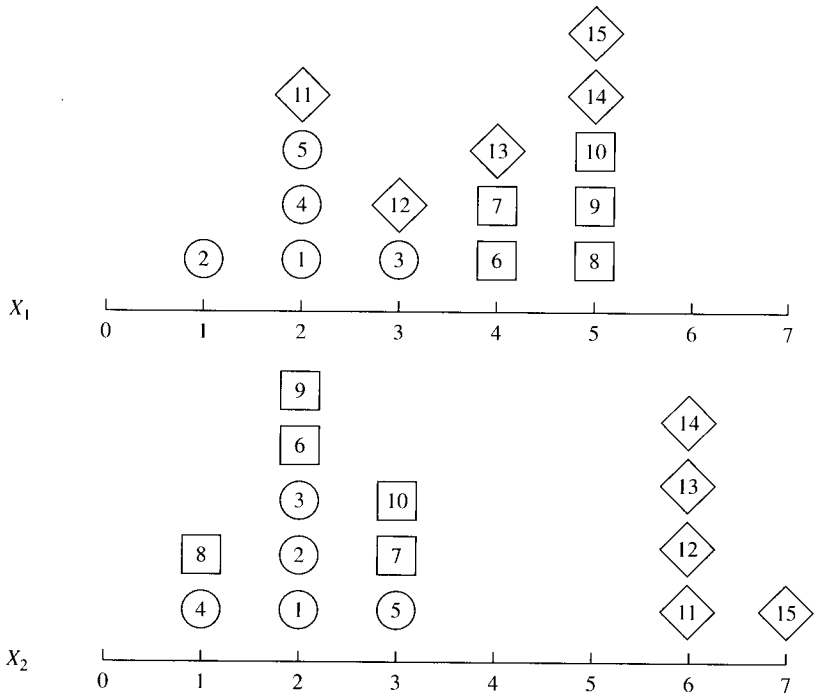
Para ilustrar esto gráficamente, la Figura 5.4 representa los tres grupos de cada una de las variables independientes de forma separada. Observando los miembros del grupo en cada variable, vemos que ninguna variable discrimina bien entre todos los grupos. Pero si ahora construimos dos funciones discriminantes simples, los resultados llegan a ser mucho más claros. Con fines ilustrativos, calculamos dos funciones discriminantes con ponderaciones de 0,0 o 1,0 para las variables. La función discriminante 1 da a X_1 una ponderación de 1,0, mientras que a X_2 se le da una ponderación de 0,0. De la misma manera, la función discriminante 2 da a X_2 una ponderación de 1,0, y a X_1 una ponderación de 0,0. Las funciones pueden establecerse matemáticamente como

$$\text{Función Discriminante 1} = 1{,}0(X_1) + 0{,}0(X_2)$$
$$\text{Función Discriminante 2} = 0{,}0(X_1) + 1{,}0(X_2)$$

Esto refleja en términos sencillos cómo el proceso del análisis discriminante estima ponderaciones para maximizar la discriminación.

Con las dos funciones, ahora podemos calcular dos puntuaciones discriminantes para cada encuestado. La Figura 5.4 también contiene un punto para cada encuestado en una representación bidimensional. La separación entre los grupos llega a ser ahora bastante evidente, y cada grupo puede ser fácilmente diferenciado. Podemos establecer valores sobre cada dimensión que definan regiones que contengan a cada grupo (por ejemplo, todos los miembros del grupo 1 están en la región menor que 3,5 sobre la dimensión 1 y menor que 4,5 sobre la dimensión 2). Cada uno de los otros grupos puede estar definido de forma similar en términos de los tamaños de las puntuaciones de sus funciones discriminantes.

A. Variables individuales



B. Representación bidimensional de funciones discriminantes

Función discriminante 1 = 1.0 * X_1 + 0 * X_2

Función discriminante 2 = 0 * X_1 + 1.0 * X_2

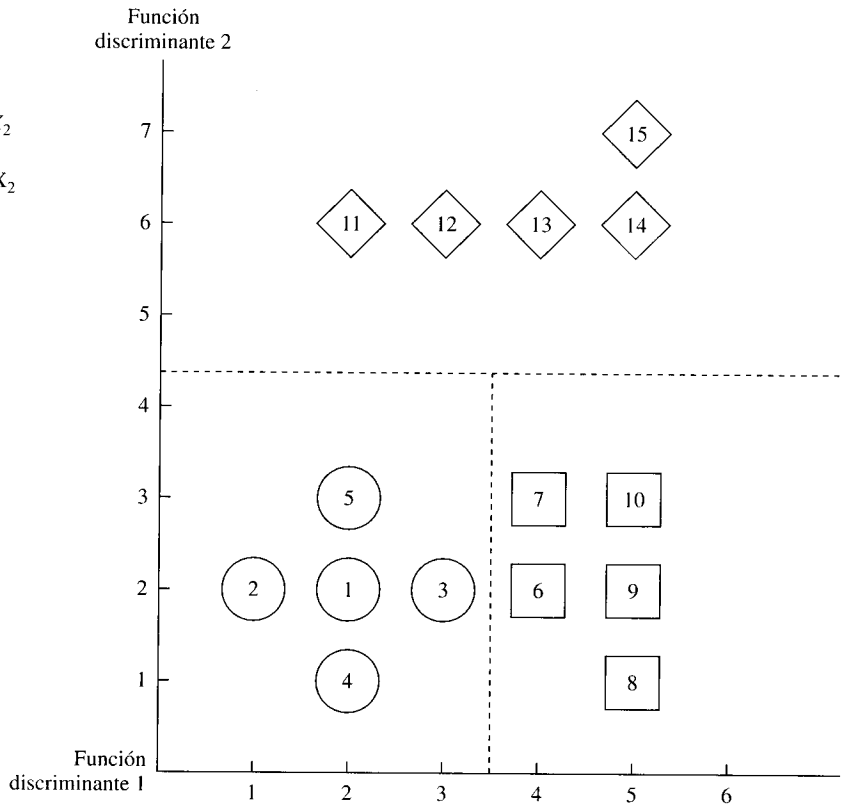


FIGURA 5.4. Representación gráfica de variables discriminantes potenciales para un análisis discriminante de tres grupos.

Además, las dos funciones discriminantes proporcionan las dimensiones de la discriminación. La primera función discriminante, la competitividad en el precio, distingue entre clientes indecisos y aquellos clientes que han decidido cambiar. Pero la competitividad en el precio no diferencia a aquellos que han decidido no cambiar. En cambio, la percepción del nivel de servicio, que define la segunda función discriminante, predice si un cliente decidirá no cambiar, frente a si un cliente está indeciso o determinado a cambiar de proveedor. El investigador puede presentar la dirección de las influencias separadas tanto de la competitividad en el precio como del nivel de servicio al tomar esta decisión.

La estimación de más de una función discriminante, cuando sea posible, proporciona al investigador tanto una mejora en la discriminación como perspectivas adicionales sobre las características y combinaciones que mejor discriminan entre los grupos. Las siguientes secciones detallan los pasos necesarios para realizar un análisis discriminante, para valorar su nivel de ajuste predictivo y, al final, para interpretar la influencia de variables independientes al llevar a cabo esa predicción.

El proceso de decisión en el análisis discriminante

Se puede considerar la aplicación del análisis discriminante desde el punto de vista de la construcción de un modelo en seis etapas, introducido en el Capítulo 1 y representado en la Figura 5.5 (etapas 1 a 3) y la Figura 5.6 (etapas 4 a 6). Como en todas las aplicaciones multivariantes, el primer paso del análisis consiste en fijar los objetivos. Después, el investigador debe centrarse en cuestiones de diseño específicas y estar seguro de que se cumplen los supuestos básicos. El análisis continúa con la derivación de la función discriminante y la determinación de si puede obtenerse o no una función estadísticamente significativa que separe los dos (o más) grupos. Los resultados discriminantes se evalúan entonces para ver la capacidad predictiva construyendo para ello una ma-

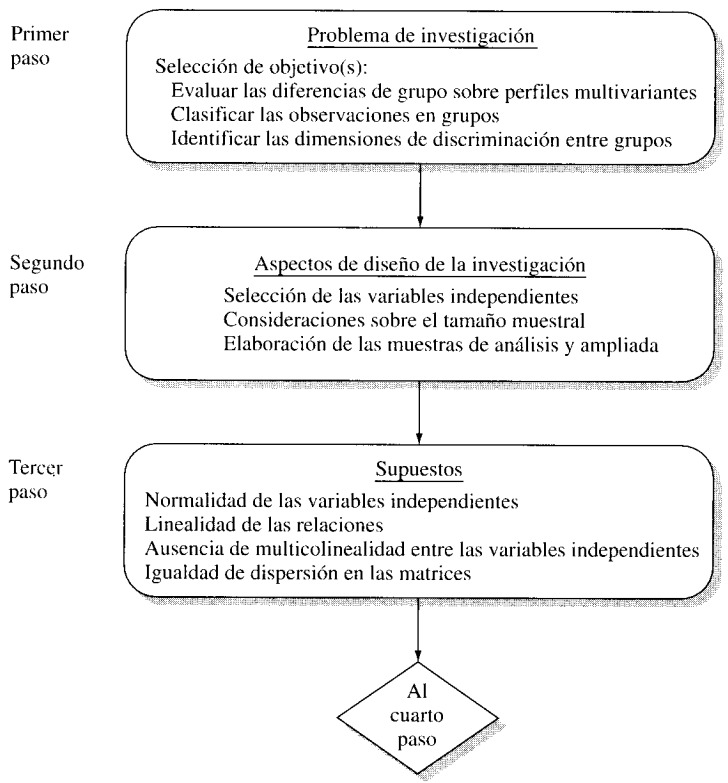


FIGURA 5.5. Pasos uno a tres en el diagrama de decisión del análisis.

triz de clasificación. El siguiente paso, la interpretación de la función discriminante, determina cuál de las variables independientes es la que más contribuye a discriminar entre los grupos. Finalmente, la función discriminante debería ser validada mediante una ampliación de la muestra. Cada uno de estos pasos se tratan en las siguientes secciones. Trataremos la regresión logística en una sección aparte tras examinar el proceso de decisión en el análisis discriminante. De esta forma, podrán aclararse las diferencias y semejanzas entre estas dos técnicas.

Primer paso: Objetivos del análisis discriminante

Un repaso de los objetivos para la utilización del análisis discriminante debería clarificar bastante su naturaleza. El análisis discriminante puede tratar cualquiera de los siguientes objetivos de investigación:

1. Determinar si existen diferencias estadísticamente significativas entre los perfiles de las puntuaciones medias sobre un conjunto de variables de dos (o más) grupos definidos a priori.
2. Determinar cuál de las variables independientes cuantifica mejor de las diferencias en los perfiles de las puntuaciones medias de dos o más grupos.
3. Establecer los procedimientos para clasificar objetos (individuos, empresas, productos, etc.), dentro de los grupos, en base a sus puntuaciones sobre un conjunto de variables independientes.
4. Establecer el número y la composición de las dimensiones de la discriminación entre los grupos formados a partir del conjunto de variables independientes.

Como se puede observar a partir de estos objetivos, el análisis discriminante es útil cuando el investigador está interesado en comprender las diferencias de los grupos o en clasificar correctamente objetos en grupos o clases. Por tanto, se puede considerar al análisis discriminante tanto un tipo de análisis de perfil como una técnica predictiva analítica. En cualquier caso, la técnica es la más apropiada cuando existe una única variable dependiente categórica y varias variables independientes escaladas métricamente. Como en el análisis de perfil, el análisis discriminante proporciona una valoración objetiva de las diferencias entre grupos sobre un conjunto de variables independientes. En este caso, el análisis discriminante es bastante similar al análisis multivariante de la varianza (véase Capítulo 6 para un tratamiento más detallado del análisis multivariante de la varianza). Para comprender las diferencias del grupo, el análisis discriminante tiene en cuenta tanto el papel de las variables independientes como las combinaciones que se construyen con estas variables que representan dimensiones de discriminación entre los grupos. Estas dimensiones son los efectos conjuntos de varias variables que trabajan unidas para diferenciar entre grupos. El uso de los métodos de estimación secuencial permite también identificar subconjuntos de variables con la mayor capacidad discriminante. Finalmente, para fines de clasificación, el análisis discriminante proporciona una base, no sólo para clasificar la muestra utilizada para estimar la función discriminante, sino también cualesquiera otras observaciones que puedan tener valores para todas las variables independientes. De esta forma, el análisis discriminante puede utilizarse para clasificar otras observaciones dentro de los grupos definidos.

Segundo paso: Diseño de la investigación mediante análisis discriminante

El éxito en la aplicación del análisis discriminante requiere tener en cuenta varias cuestiones. Estas cuestiones incluyen la selección tanto de la variable dependiente como de las independientes, el tamaño muestral necesario para la estimación de las funciones discriminantes y la división de la muestra con fines de validación.

Selección de las variables dependiente e independientes

Para aplicar el análisis discriminante, el investigador primero debe especificar qué variables van a ser independientes y qué variable va a ser dependiente. Recuerde que la variable dependiente es categórica y las variables independientes son métricas.

El investigador debe centrarse primero en la variable dependiente. El número de grupos de la variable dependiente (categorías) puede ser de dos o más, pero estos grupos deben ser mutuamente excluyentes y exhaustivos. Con esto queremos decir que cada observación debe estar colocada dentro de un grupo solamente. En algunos casos, la variable dependiente consta de dos grupos (dicotómica), como por ejemplo bueno frente a malo. En otros casos, la variable dependiente puede incluir varios grupos (multicotómica), como por ejemplo ocupaciones tales como físico, abogado o profesor.

Los ejemplos anteriores de variables categóricas constituyen verdaderas dicotomías (o multicotomías). Sin embargo, hay algunas situaciones donde el análisis discriminante es apropiado incluso aunque la variable dependiente no sea una verdadera variable categórica. Podemos tener una variable dependiente que sea ordinal o medida a intervalos que queremos utilizar como variable dependiente categórica. En tales casos, tendremos que crear una variable categórica. Por ejemplo, si tuviéramos una variable que midiera el número medio de bebidas de cola consumidas por día, y la respuesta de los individuos se basara en una escala de cero a ocho o más por día, podríamos crear una tricotomía artificial (tres grupos) simplemente designando a aquellos individuos que han consumido ninguna, una o dos bebidas al día como pequeños consumidores; a aquellos que han consumido tres, cuatro o cinco al día como consumidores medios; y a quienes consumieron seis, siete, ocho o más como grandes consumidores. Este procedimiento crearía una variable categórica de tres grupos donde el objetivo sería discriminar entre pequeños, medios y grandes consumidores de cola.

Se puede construir cualquier número de grupos categóricos artificiales. Lo más frecuente es crear dos, tres o cuatro categorías. Pero se podría establecer un número más grande de categorías si fuese necesario. Cuando se crean tres o más categorías, se presenta la posibilidad de examinar solamente los grupos extremos en un análisis discriminante de dos grupos. A este proceso se le denomina enfoque de los extremos polares.

El **enfoque de los extremos polares** compara solamente los dos grupos extremos excluyendo los grupos medios del análisis discriminante. Por ejemplo, el investigador podría examinar los pequeños y grandes consumidores de cola, y excluir a los consumidores medios. Este enfoque puede ser utilizado por el investigador en cualquier momento en que desee examinar solamente los grupos extremos. Sin embargo, el investigador puede también intentar esta aproximación cuando los resultados del análisis de la regresión no sean tan buenos como esperaba. Este proceder puede ser de ayuda porque es posible que las diferencias del grupo puedan aparecer incluso aunque los resultados de la regresión sean pobres; es decir, el enfoque de los extremos polares con análisis discriminante puede revelar diferencias que no resultan claras en un análisis de regresión de un conjunto completo de datos [6]. Tal manipulación de los datos naturalmente exige precaución al interpretar los resultados.

Después de tomarse una decisión sobre la variable dependiente, el investigador debe decidir qué variables independientes incluye en el análisis. Las variables independientes generalmente se seleccionan de dos formas. La primera implica identificar las variables tanto en la investigación previa como desde el modelo teórico que sirve de fundamento a la pregunta de la investigación. La segunda forma es intuitiva —utilizando el conocimiento del investigador y seleccionando intuitivamente las variables para las cuales no existe investigación previa o teoría, pero que lógicamente podrían relacionarse para predecir los grupos de la variable dependiente.

Tamaño muestral

El análisis discriminante es bastante sensible al ratio entre el tamaño muestral y el número de variables predictoras. Muchos estudios sugieren un ratio de 20 observaciones por cada variable pre-

dictora. Aunque este ratio puede ser difícil de conseguir en la práctica, el investigador debe tener en cuenta que los resultados podrían llegar a ser inestables a medida que el tamaño muestral disminuye en relación con el número de variables independientes. El tamaño mínimo recomendado es de cinco observaciones por variable independiente. Nótese que este ratio se aplica a todas las variables consideradas en el análisis, incluso si todas las variables consideradas no entran en la función discriminante (como en la estimación por etapas).

Además del tamaño muestral total, el investigador debe también considerar el tamaño muestral de cada grupo. Como mínimo, el tamaño del grupo más pequeño debe ser mayor que el número de variables independientes. Como una regla práctica, cada grupo debe tener al menos 20 observaciones. Pero incluso aunque todos los grupos excedan las 20 observaciones, el investigador debe también considerar los tamaños relativos de los grupos. Si los grupos varían ampliamente en tamaño, esto puede afectar a la estimación de la función discriminante y a la clasificación de las observaciones. En la etapa de clasificación, los grupos más grandes tienen una posibilidad desproporcionadamente más grande de clasificación. Si los tamaños de los grupos varían de forma importante, puede que el investigador quiera muestrear aleatoriamente desde el grupo más grande, y con ello reducir su tamaño a un nivel comparable con el grupo más pequeño.

División de la muestra

Una observación final sobre la influencia del tamaño muestral en el análisis discriminante. Como se verá más adelante, en muchas ocasiones la muestra se divide en dos submuestras; una, utilizada para la estimación de la función discriminante, y otra con fines de validación. Es esencial que cada submuestra tenga un tamaño adecuado para apoyar las conclusiones de los resultados.

Se han sugerido un conjunto de procedimientos para dividir la muestra, pero el más utilizado implica desarrollar la función discriminante con un grupo y luego probarla con un segundo grupo. El procedimiento habitual consiste en dividir aleatoriamente la muestra total de encuestados en dos grupos. Uno de estos grupos, la **muestra de análisis**, se usa para construir la función discriminante. El segundo grupo, la **ampliación de la muestra**, se usa para validar la función discriminante. Este método de validación de la función se denomina **división de la muestra o enfoque de validación cruzada** [4, 8, 15].

No se ha establecido una manera definitiva para dividir la muestra en los grupos de análisis y ampliación (o validación). El procedimiento más común es dividir el total del grupo, de tal forma que la mitad de los encuestados pertenezca a la muestra de análisis y la otra mitad a la ampliación de la muestra. No obstante, no se ha establecido ninguna regla fiable, y algunos investigadores prefieren una división 60-40 o 75-25 entre los grupos de análisis y ampliación.

Cuando se seleccionan los individuos para los grupos de análisis y validación, generalmente se sigue un proceso de muestreo estratificado proporcional. Si los grupos categóricos del análisis discriminante están igualmente representados en el total de la muestra, se selecciona un número igual de individuos. Si los grupos categóricos son desiguales, los tamaños de los grupos seleccionados para la ampliación de la muestra deben ser proporcionales a la distribución total de la muestra. Por ejemplo, si una muestra consiste en 50 hombres y 50 mujeres, la ampliación de la muestra tendría 25 hombres y 25 mujeres. Si la muestra contiene 70 mujeres y 30 hombres, entonces la ampliación de la muestra consistiría en 35 mujeres y 15 hombres.

Es necesario realizar algunos comentarios adicionales en lo referente a dividir la muestra total en los grupos de análisis y ampliación. Si el investigador va a dividir la muestra en los grupos de análisis y ampliación, la muestra debe ser suficientemente grande para realizarlo. Una vez más, no se ha establecido una regla fiable, pero parece lógico que el investigador querría al menos 100 observaciones en el total de la muestra para justificar el dividirla en dos grupos. Una solución de compromiso es que el investigador, si el tamaño muestral es demasiado pequeño como para justificar la división en los grupos de análisis y ampliación, construya la función con la muestra entera y después utilice esta función para clasificar el mismo grupo que sirvió para construirla. Este procedimiento sesga al alza la capacidad predictiva de la función, pero es ciertamente mejor que no validar la función en absoluto.

Tercer paso: Supuestos del análisis discriminante

Es deseable encontrar ciertas condiciones para la correcta aplicación del análisis discriminante. Los supuestos clave para obtener la función discriminante son el de normalidad multivariante de las variables independientes y el de estructuras (matrices) de covarianza y dispersión desconocidas (pero iguales) para los grupos, como se definió para la variable dependiente [7, 9]. Aunque existe una evidencia contradictoria sobre la sensibilidad del análisis discriminante a incumplimientos de estos supuestos, el investigador debería examinar los datos y, si los supuestos no se cumplen, debería identificar los métodos alternativos disponibles y la influencia que cabría esperar sobre los resultados. Los datos que no cumplan el supuesto de normalidad multivariante pueden causar problemas en la estimación de la función discriminante. Por ello, se sugiere que se use la regresión logística como una técnica alternativa, si es posible.

Las matrices de covarianzas distintas pueden afectar desfavorablemente al proceso de clasificación. Si los tamaños muestrales son pequeños y las matrices de covarianzas son distintas, la significación estadística del proceso de estimación se ve afectada desfavorablemente. El caso más probable es el de covarianzas distintas entre grupos de tamaño muestral adecuado, en donde las observaciones son «sobreclasificadas» dentro de los grupos con matrices de covarianzas más grandes. Este efecto puede minimizarse incrementando el tamaño muestral y también usando las matrices de covarianzas especificas de cada grupo con fines clasificatorios, pero esta aproximación obliga a la validación cruzada de los resultados discriminantes. Finalmente, en muchos de los programas estadísticos están disponibles técnicas de clasificación cuadráticas, si existen grandes diferencias entre las matrices de covarianzas de los grupos y otras soluciones no minimizan el efecto [5, 11, 13].

Otra característica de los datos que puede afectar a los resultados es la multicolinealidad entre las variables independientes. La multicolinealidad consiste en que dos o más variables independientes están altamente correlacionadas, por lo que una variable puede venir muy bien explicada o predicha por otras variables y, por ello, añadir poca capacidad explicativa al conjunto completo. Esta consideración llega a ser especialmente crítica cuando se emplean los procesos por etapas. El investigador, al interpretar la función discriminante, debe conocer el nivel de multicolinealidad y su influencia al determinar que variables entran en la solución por etapas. (Para una presentación más detallada de la multicolinealidad y su influencia en las soluciones por etapas, véase Capítulo 4).

Al igual que con alguna de las técnicas multivariantes que emplean un valor teórico, un supuesto implícito es que todas las relaciones son lineales. Las relaciones no lineales no están reflejadas en la función discriminante a menos que se realicen transformaciones específicas de la variable para representar los efectos no lineales. Finalmente, los casos atípicos pueden tener una influencia sustancial en la precisión clasificatoria de cualquier resultado del análisis discriminante. Se aconseja al investigador examinar todos los resultados por la presencia de casos atípicos y eliminarlos si fuera necesario. (Para una discusión de algunas técnicas de evaluación de los incumplimientos de los supuestos estadísticos básicos o de detección de casos atípicos, véase Capítulo 2.)

Cuarto paso: Estimación del modelo discriminante y valoración del ajuste global

Para obtener la función discriminante, el investigador debe decidir el método de estimación y determinar después el número de observaciones que se van a mantener (véase Figura 5.6). Una vez que se han estimado las funciones, puede valorarse el ajuste global del modelo de varias formas. Primero pueden calcularse las **puntuaciones discriminantes *Z***, también conocidas como **puntuaciones *Z***. La comparación de las medias de los grupos sobre las puntuaciones *Z* ofrece una me-

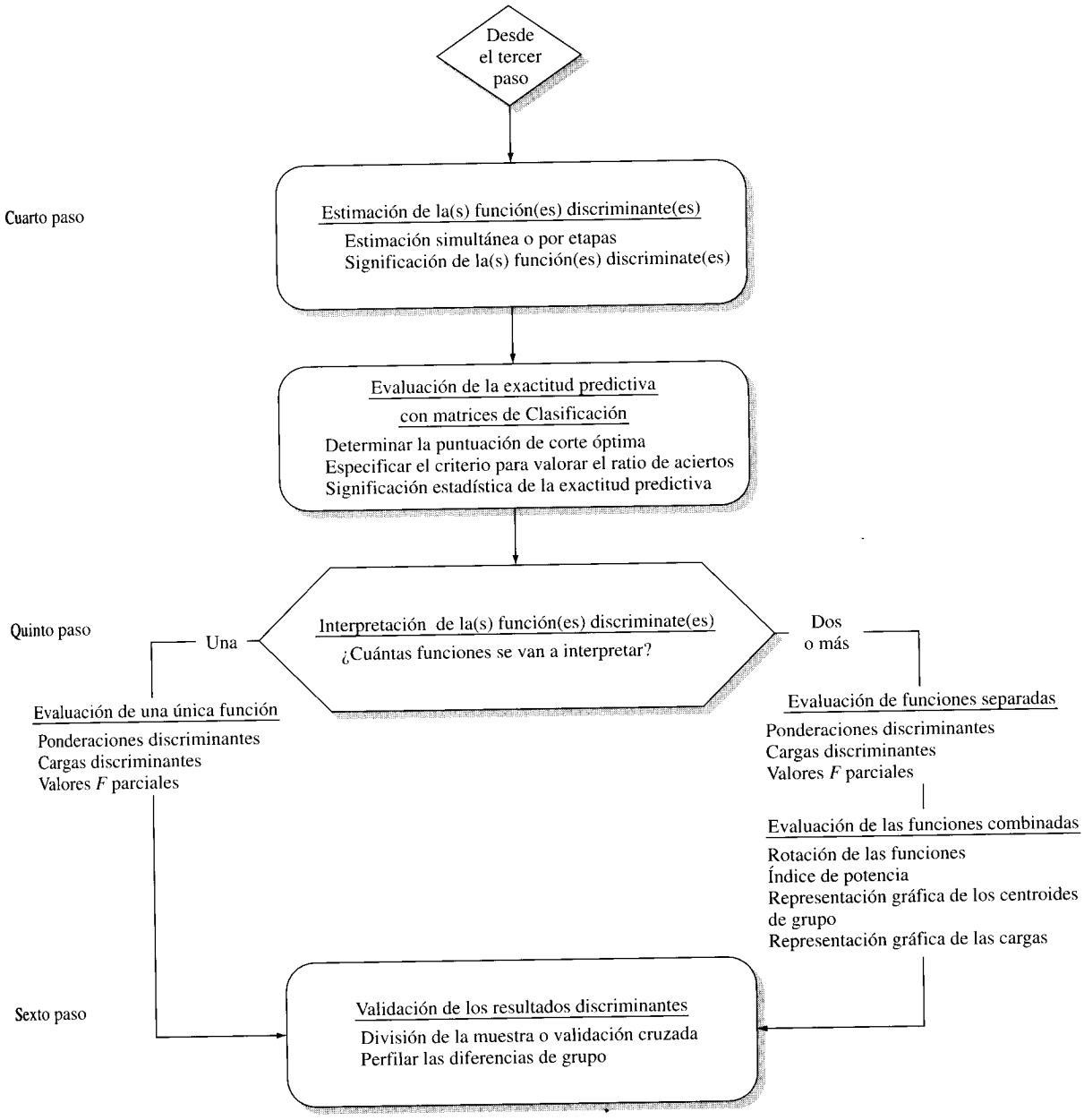


FIGURA 5.6. Pasos cuatro a seis en el diagrama de decisión del análisis discriminante.

dida de la discriminación entre grupos. La capacidad en la predicción se valora por el número de observaciones clasificadas dentro de los grupos adecuados. Se dispone de varios criterios para valorar si el proceso de clasificación alcanza significación estadística y/o práctica. Finalmente, la validación por casos puede identificar la precisión en la clasificación de cada caso y su influencia relativa sobre la estimación global del modelo.

Método de cálculo

Se pueden utilizar dos métodos de cálculo para derivar una función discriminante: el método simultáneo (directo) y el método por etapas. La **estimación simultánea** implica el cálculo de la función discriminante donde todas las variables independientes son consideradas simultáneamente. Por

ello, la(s) función(es) discriminante(s) se calculan basándose en el conjunto completo de variables independientes, sin considerar la capacidad discriminante de cada variable independiente. El método simultáneo es apropiado cuando, por razones teóricas, el investigador quiere introducir todas las variables independientes en el análisis y no está interesado en observar resultados intermedios basados solamente en las variables que discriminan mejor.

La **estimación por etapas** es una alternativa al enfoque simultáneo. Incluye las variables independientes dentro de la función discriminante de una en una, según su capacidad discriminatoria. El enfoque por etapas comienza eligiendo la variable que mejor discrimina. La variable inicial se empareja entonces con cada una de las variables independientes (de una en una), y se elige la variable que más consigue incrementar la capacidad discriminante de la función en combinación con la primera variable. La tercera y posteriores variables se seleccionan de una manera similar. Mientras se incluyen variables adicionales, algunas variables seleccionadas previamente pueden ser eliminadas si la información que contienen sobre las diferencias del grupo está contenida en alguna combinación de otras variables incluidas en posteriores etapas. Al final, o bien todas las variables habrán sido incluidas en la función, o se habrá considerado que las variables excluidas no contribuyen significativamente a una mejor discriminación.

El método por etapas es útil cuando el investigador quiere considerar un número relativamente grande de variables independientes para incluir en la función. Seleccionando secuencialmente la siguiente variable que mejor discrimina en cada paso, se eliminan las variables que no son útiles para discriminar entre los grupos y se identifica un conjunto reducido de variables. El conjunto reducido es generalmente tan bueno como, y algunas veces mejor que, el conjunto completo de variables. Pero el investigador debería darse cuenta de que la estimación por etapas puede llegar a ser menos estable y generalizable en tanto que el ratio del tamaño muestral respecto a las variables independientes se reduce por debajo de las 20 observaciones por variable independiente. Es especialmente importante en estas situaciones validar los resultados de tantas formas como sea posible.

Significación estadística

Después de calcularse la función discriminante, el investigador debe valorar el nivel de significación. Se dispone de varios criterios estadísticos. Las medidas del lambda de Wilks, la traza de Hotelling y el criterio de Pillai evalúan la significación estadística de la capacidad discriminatoria de la función(es) discriminante(s). La mayor raíz característica de Roy evalúa solamente la primera función discriminante. Para un tratamiento más detallado de las ventajas y desventajas de cada criterio, se remite al lector a la discusión del contraste de significación en el análisis multivariante de la varianza del Capítulo 6.

Si se utiliza un método por etapas para estimar la función discriminante, son más apropiadas las medidas D^2 de Mahalanobis y V de Rao. Ambas son medidas de distancia generalizada. El procedimiento de la D^2 de Mahalanobis se basa en la distancia euclídea al cuadrado generalizada que se adecúa a varianzas desiguales. La principal ventaja de este procedimiento es que se calcula en el espacio original de las variables predictoras, en lugar de alguna otra versión obsoleta utilizada en otras medidas. El procedimiento D^2 de Mahalanobis llega a ser particularmente crítico a medida que el número de variables predictoras se incrementa dado que no se da ninguna reducción en la dimensionalidad. La pérdida de dimensionalidad causa una pérdida de información, porque hace disminuir la variabilidad de las variables independientes. En general, el procedimiento de la D^2 de Mahalanobis es el preferido cuando se está interesado en aprovechar al máximo la información disponible. El procedimiento D^2 de Mahalanobis lleva a cabo un análisis discriminante por etapas similar al análisis de regresión por etapas. Este procedimiento por etapas esta diseñado para obtener el mejor modelo de una variable, seguido por el mejor modelo de dos variables, y así sucesivamente hasta que ninguna otra variable cumpla la regla de selección deseada. La regla de selección en este procedimiento es maximizar la distancia D^2 de Mahanalobis entre los grupos. Tanto el método por etapas como el simultáneo están disponibles en los principales programas estadísticos.

El criterio convencional de 0,05 o superior se utiliza a menudo. Muchas investigadores creen que si la función no es significativa a ese nivel o más, existe poca justificación para seguir adelante.

Algunos investigadores, sin embargo, no están de acuerdo. Su regla de decisión para continuar hasta un nivel de significación mayor (0,10 o más, por ejemplo) resulta de comparar el coste de la información frente a su valor. Si son aceptables mayores niveles de riesgo por incluir resultados no significativos (por ejemplo, niveles de significación mayores a 0,5), se pueden mantener funciones discriminantes que son significativas al nivel del 0,2 o incluso al 0,3.

Si el número de grupos es de tres o más, entonces el investigador debe decidir no solamente si la discriminación entre el total de los grupos es estadísticamente significativa, sino también si cada una de las funciones discriminantes estimadas es estadísticamente significativa. Como se discutió anteriormente, el análisis discriminante estima una función discriminante menos que grupos existentes. Si se analizan tres grupos, se estimarán dos funciones discriminantes. Para cuatro grupos, se estimarán tres funciones, y así sucesivamente. Todos los programas de computador proporcionan al investigador la información necesaria para averiguar el número de funciones necesarias para obtener significación estadística, sin incluir funciones discriminantes que no incrementen la capacidad discriminatoria significativamente. Si se consideran una o más funciones que no son estadísticamente significativas, el modelo discriminante debería reestimarse con el número de funciones que se hayan obtenido, limitado por el número de funciones significativas. De esta manera, la valoración de la precisión en la predicción y la interpretación de las funciones discriminantes estarán basadas solamente en funciones significativas.

Valoración del ajuste global

Una vez que se han identificado las funciones discriminantes significativas, la atención se desplaza a averiguar el ajuste global de la(s) función(es) discriminante(s) considerada(s). Esta valoración conlleva tres tareas: calcular la puntuación *Z* discriminante para cada observación, evaluar diferencias de grupo sobre las puntuaciones *Z* discriminantes y valorar la precisión en la predicción de pertenencia al grupo.

Cálculo de las puntuaciones Z discriminantes

Definidas las funciones discriminantes, se han establecido las bases para el cálculo de las puntuaciones *Z* discriminantes. Como se discutió más arriba, la puntuación *Z* discriminante de cualquier función discriminante puede calcularse para cada observación mediante la siguiente fórmula:

$$Z_{jk} = a + W_1X_{1k} + W_2X_{2k} + \cdots + W_nX_{nk}$$

donde

Z_{jk} = puntuación *Z* discriminante de la función discriminante *j* para el objeto *k*
a = constante
W_i = ponderación discriminante para la variable independiente *i*
X_{ik} = variable independiente *i* para el objeto *k*

Esta puntuación, una medida métrica, ofrece unas medias directas para comparar observaciones para cada función. Las observaciones con puntuaciones *Z* similares se suponen mas parecidas sobre las variables que constituyen esta función que aquellas con puntuaciones dispares. Estas son versiones de la función discriminante que emplean valores y ponderaciones estandarizadas o no estandarizadas. La versión estandarizada es más útil en la interpretación, pero la versión no estandarizada es más fácil de utilizar en el cálculo de la puntuación *Z* discriminante.

Debemos darnos cuenta de que la función discriminantes difiere de la **función de clasificación**, también conocida como la **función discriminante lineal de Fisher**. Las funciones de clasificación, una para cada grupo, pueden utilizarse al clasificar observaciones. En este método de clasificación, unos valores de la observación para las variables independientes se incluyen en las funciones de clasificación y se calcula una puntuación de clasificación para cada grupo para esa observación.

La observación se clasifica entonces en el grupo con la mayor puntuación de clasificación. Utilizamos la función discriminante como el medio de clasificar porque ofrece una representación resumida y simple de cada función discriminante, simplificando el proceso de interpretación y la valoración de la contribución de las variables independientes.

Evaluación de diferencias entre grupos

Una forma de valorar el ajuste global del modelo es determinar la magnitud de las diferencias entre los miembros de cada grupo en términos de las puntuaciones Z discriminantes. Una medida resumen de las diferencias entre grupos es una comparación de los **centroides** de grupo, la puntuación Z discriminante media para todos los miembros del grupo. Una medida de éxito del análisis discriminante es su capacidad para definir funciones discriminantes que den lugar a centroides de grupo significativamente diferentes. Las diferencias entre centroides se miden en términos de la medida D^2 de Mahalanobis, para la cual se dispone de contrastes que determinan si las diferencias son significativamente distintas. El investigador debería asegurarse de que incluso con funciones discriminantes significativas existen diferencias significativas entre cada uno de los grupos.

Los centroides de grupo entre cada función discriminante pueden también dibujarse para mostrar los resultados desde una perpectiva global. Normalmente los gráficos se realizan para las primeras dos o tres funciones discriminantes (suponiendo que son estadísticamente significativas y funciones predictivas válidas). Los valores para cada grupo muestran su posición en un espacio discriminante reducido (llamado así porque no todas las funciones y por tanto no toda la varianza está representada). El investigador puede ver las diferencias entre los grupos sobre cada función; sin embargo, la inspección visual no explica totalmente lo que son estas diferencias. Se pueden dibujar círculos que incluyan la distribución de las observaciones alrededor de sus respectivos centroides para aclarar más las diferencias entre grupos, pero este procedimiento va más allá de los objetivos de este texto (véase Dillon y Golstein [3]).

Valorando la exactitud en la predicción de pertenencia al grupo

Dado que la variable dependiente es no métrica, no es posible utilizar una medida como el R^2, como se hace en la regresión múltiple, para valorar la exactitud predictiva. En su lugar, cada observación debe valorarse como si fuera correctamente clasificada. Al hacer esto, deben realizarse una serie de consideraciones: la razón de ser práctica y estadística para elaborar matrices de clasificación, la determinación de la puntuación de corte, la construcción de matrices de clasificación y los estándares para valorar la exactitud clasificatoria.

Por qué se elaboran matrices de clasificación Los contrastes estadísticos para valorar la significación de las funciones discriminantes no informan sobre lo correctamente que predice la función. Por ejemplo, suponga dos grupos que son significativamente diferentes por encima del nivel del 0,01. Con tamaños de la muestra suficientemente grandes, las medias de los grupos (centroides) podrían ser virtualmente idénticas y todavía tendríamos significación estadística. En resumen, estos contrastes adolecen de los mismos inconvenientes que los contrastes clásicos de hipótesis. Por esto, el nivel de significación de estos estadísticos es una indicación muy pobre de la capacidad de la función para discriminar entre los dos grupos. Para determinar la capacidad predictiva de una función discriminante, el investigador debe construir matrices de clasificación.

Para aclarar mejor la utilidad del procedimiento de las matrices de clasificación, lo relacionaremos con el concepto de R^2 en el análisis de regresión. La mayoría de nosotros hemos leído artículos académicos en los que el autor ha encontrado relaciones estadísticamente significativas, sin embargo ha explicado solamente el 10 por ciento (o menos) de la varianza ($R^2 = 0{,}10$). Generalmente este R^2 es significativamente distinto de cero simplemente porque el tamaño muestral es grande. En el análisis discriminante múltiple, el **ratio de aciertos** (porcentaje correctamente clasificado) es análogo al R^2 de la regresión. El ratio de aciertos revela lo correctamente que la función discri-

minante clasificó los objetos; el R^2 indica cuánta varianza explicó la ecuación de regresión. El contraste F para la significación estadística del R^2 es, por tanto, análogo al contraste chi-cuadrado (o D^2) de significación en el análisis discriminante. Claramente, con un tamaño muestral suficientemente grande en el análisis discriminante, podríamos tener una diferencia estadísticamente significativa entre los dos (o más) grupos y sin embargo clasificar correctamente solamente el 53 por ciento (cuando la probabilidad es el 50 por ciento con tamaños de grupo iguales) [14].

Determinación de la puntuación de corte Si los contrastes estadísticos indican que la función discrimina significativamente, es usual elaborar matrices de clasificación para proporcionar una valoración más precisa de la capacidad discriminatoria de la función. Sin embargo, antes de que pueda ser construida una matriz de clasificación el investigador debe determinar la puntuación de corte. La **puntuación de corte** es el criterio (puntuación) frente al cual cada puntuación discriminante individual es comparada para determinar dentro de qué grupo debe ser clasificado cada objeto.

Al construir las matrices de clasificación, el investigador querrá determinar la **puntuación de corte óptima** (también llamada valor Z crítico). La puntuación de corte óptima diferirá dependiendo de si los tamaños de los grupos son iguales o distintos. Si los grupos son de igual tamaño, la puntuación de corte óptima estará a mitad de camino entre los centroides de los dos grupos. El punto de corte para dos grupos de igual tamaño se define por tanto como:

$$Z_{CE} = \frac{Z_A + Z_B}{2}$$

donde

Z_{CE} = valor de la puntuación de corte crítica para grupos de igual tamaño
Z_A = centroide del grupo A
Z_B = centroide del grupo B

Especificación de las probabilidades de clasificación para tamaños de grupo distintos.

Para calcular correctamente la puntuación de corte cuando los tamaños de grupo son distintos, el investigador debe también determinar si los tamaños de los grupos observados reflejan las proporciones poblacionales reales o si los tamaños de los grupos poblacionales se deben considerar iguales. El supuesto por defecto es que las probabilidades sean iguales: en otras palabras, se supone que cada grupo tiene una misma probabilidad de ocurrir incluso aunque los tamaños de los grupos en la muestra sean distintos. Si el investigador no está seguro de si las proporciones observadas en la muestra son representativas de las proporciones de la población, la solución conservadora es suponer la igualdad en las probabilidades. Sin embargo, si la muestra está tomada aleatoriamente de la población por lo que los grupos sí representan las proporciones poblacionales en cada grupo, entonces la mejor estimación de las probabilidades anteriores no es la de igualdad, sino la de las proporciones muestrales. La influencia de especificar las anteriores probabilidades como iguales a las proporciones muestrales varía según la diferencia que exista entre las proporciones muestrales y las proporciones poblacionales. Pero el investigador debería determinar las probabilidades en todos los análisis (bien como iguales o bien basadas en los tamaños muestrales) para asegurar que los supuestos adecuados están presentes en los procesos de clasificación.

Determinación de la puntuación de corte para grupos de tamaño desigual

Si los grupos no son de igual tamaño y se supone que son representativos de las proporciones de la población, una media ponderada de los centroides de los grupos proporcionará una puntuación de corte óptima para una función discriminante. Se calcula como sigue:

$$Z_{CU} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

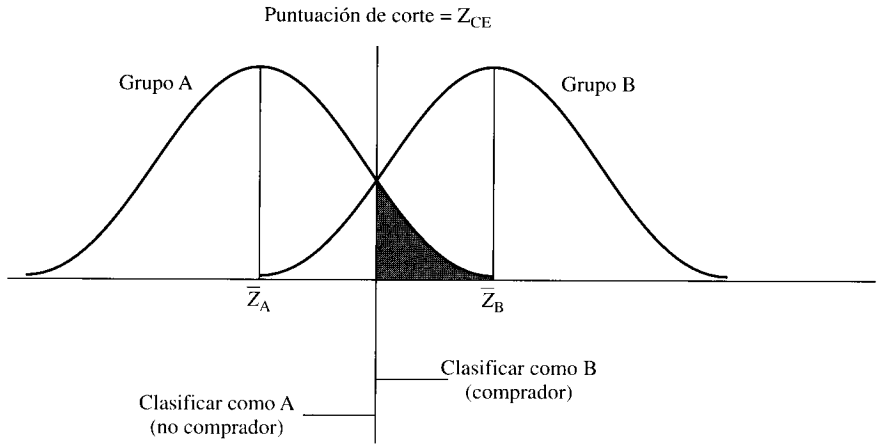


FIGURA 5.7. Puntuación de corte óptima con tamaños muestrales iguales.

donde

Z_{CU} = valor de la puntuación de corte crítica para tamaños de grupo distintos
N_A = número del grupo A
N_B = número del grupo B
$\bar{Z}_A$ = centroide del grupo A
$\bar{Z}_B$ = centroide del grupo B

Las dos fórmulas con las que se calcula la puntuación de corte óptima suponen que las distribuciones están distribuidas normalmente y se conocen las estructuras de dispersión de los grupos.

Las Figuras 5.7 y 5.8 ilustran el concepto de una puntuación de corte óptima para grupos iguales y distintos. Se muestran tanto las puntuaciones de corte ponderadas como no ponderadas. Es evidente que si el grupo A es mucho más pequeño que el grupo B, el punto de corte óptimo estará más cercano al centroide del grupo A que al centroide del grupo B. Además, si se utilizase una puntuación de corte no ponderada, ninguno de los objetos del grupo A estaría mal clasificado, pero una parte importante de los que están en el grupo B sí lo estaría.

Costes de la clasificación errónea La puntuación de corte óptima también debe tener en cuenta el coste de clasificar de forma incorrecta un objeto dentro de un grupo erróneo. Si los costes de clasificar incorrectamente a un individuo son aproximadamente iguales para todos los gru-

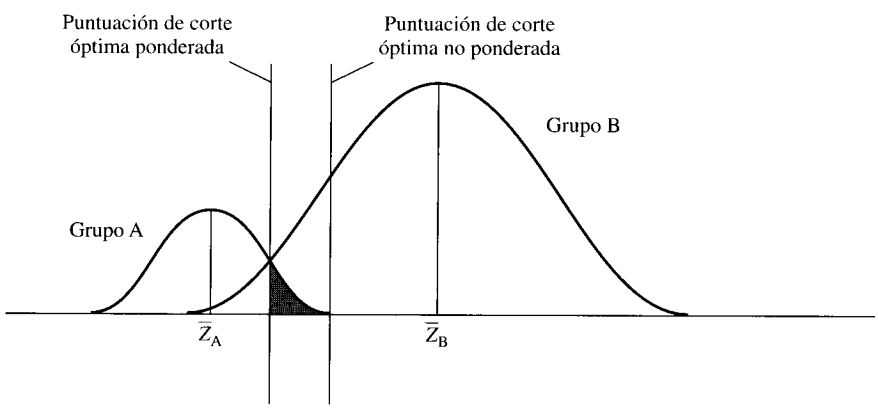


FIGURA 5.8. Puntuación de corte óptima con tamaños muestrales distintos.

pos, la puntuación de corte óptima será aquella que clasifique mal el menor número de objetos entre todos los grupos. Si los costes de la clasificación errónea son distintos, la puntuación de corte óptima será aquella que minimice dichos costes. Se consideran enfoques más sofisticados para determinar las puntuaciones de corte en Dillon y Golstein [3] y Huberty [12]. Estos enfoques están basados en un modelo estadístico bayesiano y son apropiados cuando los costes de clasificar mal en ciertos grupos son muy altos, cuando los grupos son de tamaños enormemente diferentes o cuando se quiere aprovechar un conocimiento a priori de las probabilidades de la pertenencia a un grupo.

En la práctica, al calcular la puntuación de corte generalmente no es necesario incluir las medidas de la variable primaria para cada individuo dentro de la función discriminante, y obtener la puntuación discriminante para cada persona para utilizarlo en el cálculo de Z_A y Z_B (centroides de los grupos A y B). En muchos casos el programa de computador proporciona las puntuaciones discriminantes, al igual que Z_A y Z_B, como output habitual. Cuando el investigador tiene los centroides de los grupos y los tamaños muestrales, debe únicamente sustituir los valores en la fórmula adecuada para obtener la puntuación de corte óptima.

Construcción de las matrices de clasificación Para validar la función discriminante por medio de matrices de clasificación, la muestra debe dividirse aleatoriamente en dos grupos. Uno de los grupos (la muestra de análisis) se utiliza para calcular la función discriminante. El otro grupo (la muestra de validación o ampliación de la muestra) se usa para la elaboración de la matriz de clasificación. El proceso consiste en multiplicar las ponderaciones generadas por la muestra de análisis por las medidas de la variable primaria de la ampliación de la muestra. Después, las puntuaciones discriminantes individuales para la ampliación de la muestra se comparan con el valor de la puntuación de corte crítica y se clasifican de la siguiente forma:

Clasificar a un individuo dentro del grupo A si $Z_n < Z_{ct}$.

o

Clasificar a un individuo dentro del grupo B si $Z_n > Z_{ct}$.

donde

Z_n = puntuación Z discriminante para el individuo n-ésimo
Z_{ct} = valor de la puntuación de corte crítica

Los resultados del proceso de clasificación se presentan de forma matricial, como se muestra en la Tabla 5.4. Los elementos de la diagonal de la matriz representan el número de individuos correctamente clasificados. Los números fuera de la diagonal representan las clasificaciones incorrectas. Los números de la columna denominada «tamaño del grupo real» representan el número de individuos que realmente hay en uno de los dos grupos. Los números que están al final de las columnas representan el número de individuos asignados a los grupos por la función discriminante. El porcentaje correctamente clasificado en cada grupo aparece en el lado derecho de la matriz y el porcentaje total correctamente clasificado, también conocido como ratio de aciertos, aparece al final.

TABLA 5.4. Matriz de clasificación para el análisis discriminante de dos grupos

	Grupo predicho			
Grupo real	1	2	*Tamaño del grupo real*	*Porcentaje correctamente clasificado*
1	22	3	25	88
2	5	20	25	80
Tamaño del grupo predicho	27	23	50	84[a]

[a]Porcentaje correctamente clasificado = (Número correctamente clasificado/Número total de observaciones) × 100
= [(22 + 20)/50] × 100
= 84%

En nuestro ejemplo, el número de individuos correctamente asignado en el grupo 1 es de 22, mientras que 3 miembros del grupo 1 fueron asignados incorrectamente al grupo 2. De modo similar, el número de clasificaciones correctas en el grupo 2 es de 20 y el número de asignaciones incorrectas en el grupo 1 es de 5. Con ello el porcentaje de precisión clasificatoria de la función discriminante para los verdaderos grupos 1 y 2 sería del 88 y del 80 por ciento, respectivamente. La precisión total de la clasificación (ratio de aciertos) sería del 84 por ciento.

Un último tema referido al procedimiento de clasificación es el contraste t. Se dispone de un contraste t para determinar el nivel de significación para la precisión clasificatoria. La expresión para un análisis de dos grupos (con tamaño muestral igual) es

$$t = \frac{p - 0{,}5}{\sqrt{\frac{0{,}5(1{,}0 - 0{,}5)}{N}}}$$

donde

p = proporción clasificada correctamente
$N \triangleq$ tamaño muestral

Esta expresión puede adaptarse para su uso cuando existen más grupos y para tamaños muestrales distintos.

Medición de la capacidad predictiva mediante la aleatoriedad Como se mencionó anteriormente, la capacidad predictiva de la función discriminante se mide con el ratio de aciertos, el cual se obtiene de la matriz de clasificación. El investigador podría preguntarse sobre qué se considera un nivel aceptable de capacidad predictiva para una función discriminante. Por ejemplo, ¿es el 60 por ciento un nivel aceptable o debería esperarse un 80 o un 90 por ciento de capacidad predictiva? Para responder a esta pregunta, el investigador debe determinar primero el porcentaje que podría ser clasificado correctamente de forma aleatoria (sin la ayuda de la función discriminante).

Determinación del criterio basado en la aleatoriedad

Cuando los tamaños muestrales son iguales, la determinación de la clasificación aleatoria es bastante simple; se obtiene dividiendo 1 por el número de grupos. La fórmula es C = 1/(número de grupos). Por ejemplo, en una función de dos grupos la probabilidad sería de 0,50; para una función de tres grupos la probabilidad sería de 0,33, y así sucesivamente.

El establecimiento de la clasificación aleatoria en situaciones donde los tamaños de los grupos son distintos es algo más complicado. Supongamos que tenemos una muestra en la que 75 sujetos pertenecen a un grupo y 25 al otro. Podríamos asignar arbitrariamente a todos los sujetos al grupo más grande y conseguir un 75 por ciento de capacidad predictiva sin la ayuda de la función discriminante. Se podría concluir que a menos que la función discriminante lograse una capacidad predictiva de más del 75 por ciento, no se tendría en consideración ya que no nos ayudaría a mejorar nuestra capacidad predictiva.

Determinar la clasificación aleatoria basándose en el tamaño muestral del grupo más grande se conoce como **criterio de máxima aleatoriedad**. Se determina calculando el porcentaje de la muestra completa representado por el más grande de los dos (o más) grupos. Por ejemplo, si los tamaños de los grupos son 65 y 35, el criterio de máxima aleatoriedad es 65 por ciento de clasificaciones correctas. Por tanto, si el ratio de aciertos para la función discriminante no excedió el 65 por ciento, entonces no nos ayudaría a predecir basarnos en este criterio.

El criterio de máxima aleatoriedad debería utilizarse cuando el único objetivo del análisis discriminante es maximizar el porcentaje clasificado correctamente [14]. Sin embargo, son escasas las situaciones en que sólo nos importa la maximización del porcentaje correctamente clasificado. Habitualmente el investigador utiliza el análisis discriminante para identificar correctamente a los miembros de los grupos. En casos donde los tamaños muestrales son distintos y el investigador quie-

re clasificar a los miembros de los grupos, la función discriminante desafía lo extraño clasificando a un sujeto en un grupo(s) más pequeños(s). Pero el criterio de aleatoriedad no tiene este hecho en cuenta [14]. Por lo que otro criterio de aleatoriedad —el **criterio de aleatoriedad proporcional**— debe emplearse en tales situaciones.

El criterio de aleatoriedad proporcional debe emplearse cuando los tamaños de los grupos son distintos y el investigador desea identificar adecuadamente a los miembros de los dos (o más) grupos. La fórmula para este criterio es

$$C_{\text{PRO}} = p^2 + (1 - p)^2$$

donde

p = proporción de individuos del grupo 1
$1 - p$ = proporción de individuos del grupo 2

Empleando los tamaños de los grupos de nuestro ejemplo anterior (75 y 25), vemos que el criterio de aleatoriedad proporcional sería del 62,5 por ciento frente al 75 por ciento. Por lo tanto, en este ejemplo, una capacidad predictiva del 75 por ciento sería aceptable porque está por encima del 62,5 por ciento del criterio de aleatoriedad proporcional.

Estos criterios de aleatoriedad son útiles sólo cuando se calculan con ampliación de la muestra (enfoque de división de la muestra). Si los individuos utilizados para calcular la función discriminante son los que están siendo clasificados, el resultado estará sesgado al alza en su capacidad predictiva. En tales casos, estos dos criterios tendrían que ser ajustados a su vez al alza para tener en cuenta este sesgo.

Comparación del ratio de aciertos con el criterio de aleatoriedad

La cuestión de la precisión en la clasificación es crucial. Si el porcentaje de clasificaciones correctas es significativamente más grande que el que cabría esperar de forma aleatoria, se puede llevar a cabo un ejercicio de interpretación de las funciones discriminantes con la finalidad de elaborar perfiles de grupo. Sin embargo, si la precisión clasificatoria no es más grande que lo que se podría esperar aleatoriamente, sean cuales sean las diferencias estructurales que parezcan existir, apenas se aportaría nada a la interpretación; es decir, las diferencias en los perfiles de las puntuaciones no proporcionan una información significativa para identificar la pertenencia a un grupo.

Entonces, la cuestión es, ¿cómo ha de ser la precisión en la clasificación en relación al criterio de aleatoriedad? Por ejemplo, si la probabilidad es del 50 por ciento (dos grupos, mismo tamaño muestral) ¿una precisión clasificatoria (predictiva) del 60 por ciento justifica el pasar a la etapa de interpretación? No existen directrices generales para responder a esta pregunta. En última instancia, la decisión depende de los costes relacionados con el valor de la información. Si los costes asociados con un 60 por ciento de capacidad predictiva (en relación con un 50 por ciento de forma aleatoria) son más grandes que el valor que se deriva de los resultados, no hay justificación para pasar a la interpretación. Si el valor es alto en relación a los costes, el 60 por ciento de precisión justificaría pasar a la etapa de interpretación.

El argumento de costes versus valor ofrece escasa ayuda al investigador de datos poco experimentado; aún así se sugiere el siguiente criterio: la precisión clasificatoria debería ser por lo menos un cuarto mayor que aquella obtenida por aleatoriedad. Por ejemplo, si la precisión aleatoria es del 50 por ciento, la precisión clasificatoria debe ser del 62,5 por ciento (62,5% = 1,25 × 50%). Si la precisión aleatoria es del 30 por ciento, la precisión clasificatoria debe ser del 37,5 por ciento. Este criterio proporciona solamente una estimación burda del nivel aceptable de capacidad predictiva. El criterio es fácil de aplicar con grupos de igual tamaño. Con grupos de tamaño diferente, se alcanza una cota superior cuando se utiliza el modelo de aleatoriedad máxima para determinar la precisión aleatoria. Esto no representa un gran problema sin embargo, ya que en la mayoría de las circunstancias el modelo de aleatoriedad máxima no se utilizaría con grupos de distinto tamaño.

Medidas de precisión clasificatoria fundamentadas estadísticamente relacionadas con la aleatoriedad

Un contraste estadístico para contrastar la capacidad discriminatoria de la matriz de clasificación cuando se compara con un modelo de aleatoriedad es el **estadístico *Q* de Press**. Esta medida sencilla compara el número de clasificaciones correctas con el tamaño muestral total y el número de grupos. Se compara el valor hallado con un valor crítico (el valor de la chi-cuadrado para un grado de libertad al nivel de confianza deseado). Si éste excede el valor crítico, la matriz de clasificación puede considerarse estadísticamente mejor que la aleatoriedad. El estadístico Q se calcula mediante la siguiente fórmula:

$$Q \text{ de Press} = \frac{[N - (nK)]^2}{N(K - 1)}$$

donde

N = tamaño muestral total
n = número de observaciones correctamente clasificadas
K = número de grupos

Por ejemplo, en la Tabla 5.4, el estadístico Q se calcularía con una muestra total de $N = 50$, $n = 42$ observaciones clasificadas correctamente y $K = 2$ grupos. El estadístico calculado sería

$$Q \text{ de Press} = \frac{[50 - (42 \times 2)]^2}{50(2 - 1)} = 23{,}12$$

El valor crítico a un nivel de significación de 0,01 es de 6,63. Por ello concluiríamos que en el ejemplo, las predicciones fueron significativamente mejores que las obtenidas aleatoriamente, lo cual daría una tasa correcta de clasificación del 50 por ciento. Este sencillo contraste es sensible al tamaño muestral, de tal forma que con muestras grandes es más probable que resulte significativo que con tamaños muestrales pequeños para la misma tasa de clasificación. Por ejemplo, si el tamaño muestral se incrementa hasta 100 en el ejemplo y la tasa de clasificación permanece en el 84 por ciento, el estadístico Q se incrementa a 46,24. Sin embargo, se debe tener cuidado en derivar conclusiones basadas solamente en este estadístico ya que según el tamaño muestral crece, una tasa de clasificación más baja seguiría considerándose significativa.

Diagnosis mediante casos

La última forma de evaluar el ajuste del modelo es examinar los resultados predictivos basándonos en un análisis caso por caso. Similar al análisis de los residuos en la regresión múltiple, se trata de entender qué observaciones (1) han sido mal clasificadas, y (2) no son representativas del resto de los miembros del grupo. Aunque la matriz de clasificación ofrece la exactitud de la clasificación global, no detalla los resultados para casos individuales. Además, incluso si pudiéramos saber qué casos están correctamente clasificados y cuáles no, seguiríamos necesitando una medida de similitud de la observación con respecto al resto del grupo.

Clasificación errónea de casos individuales

Al analizar los residuos en el modelo de regresión múltiple, una decisión importante es la relativa al establecimiento del nivel de los residuos considerado sustantivo y digno de atención. En el análisis discriminante este asunto es algo más sencillo, pues una observación está bien clasificada o mal clasificada. La finalidad de identificar y analizar las observaciones mal clasificadas es identi-

ficar cualesquiera características de estas observaciones que podrían incorporarse en el análisis discriminante para mejorar su capacidad predictiva. Este análisis puede adoptar la forma de perfilar los casos mal clasificados, bien sobre las variables independientes o sobre otras variables no incluidas en el modelo.

Examinando estos casos sobre las variables independientes se pueden identificar tendencias no lineales u otras relaciones o atributos que conducen a la clasificación errónea. Inspeccionar otras variables por sus diferencias entre estos casos sería el primer paso para su inclusión posible en el análisis discriminante. Aunque no existe un análisis preestablecido, como el encontrado en la regresión mútiple, se recomienda al investigador que evalúe estos casos mal clasificados desde diversas perspectivas en un intento por descubrir los rasgos únicos con que cuentan en comparación con otros miembros del grupo.

El investigador puede también realizar alguna valoración sobre la similitud de una observación con los otros miembros del grupo evaluando la distancia D^2 de Mahalanobis de la observación a la centroide del grupo. Las observaciones cercanas al centroide se supone que son representativas del grupo en una mayor medida. En el análisis gráfico de las observaciones el investigador puede identificar observaciones atípicas y realizar alguna valoración sobre su influencia en los resultados. Por ejemplo, en una situación con dos grupos, un miembro del grupo A puede tener una gran distancia D^2 de Mahalanobis, indicando así que es poco representativo del grupo. Sin embargo si esa distancia está lejos de la centroide del grupo B, incrementaría realmente la aleatoriedad de una clasificación correcta, incluso siendo la menos representativa del grupo. Una menor distancia que sitúe una observación entre dos centroides probablemente contaría con una menor probabilidad de clasificación correcta, incluso entando más cerca de su centroide de grupo que en la situación anterior.

Una representación gráfica de las observaciones es otra forma de examinar las características de las observaciones, concretamente de las mal clasificadas. Una forma habitual de proceder es representar la observación basada en sus puntuaciones Z discriminantes y reflejar el solapamiento entre grupos y los casos mal clasificados. Si se mantienen dos o más funciones, los puntos de corte óptimos pueden también representarse dando lugar a lo que se conoce como **mapa territorial** que delimita las regiones correspondientes a cada grupo. Representar los casos individuales junto con los centroides de grupo, como se señaló más arriba, muestra no sólo las caracteríticas generales del grupo representadas por los centroides, sino también la variación entre los miembros del grupo. Esto es análogo a las áreas definidas en el ejemplo con tres grupos del comienzo de este capítulo, en el que las puntuaciones de corte sobre ambas funciones definían áreas correspondientes a predicciones de clasificación para cada grupo.

Resumen

La fase de estimación y valoración tiene grandes parecidos con las otras técnicas de dependencia, permitiendo bien un proceso de estimación directo o bien por etapas, y un análisis de la capacidad predictiva global y por casos. El investigador debería prestar considerable atención a estos aspectos para evitar un uso inadecuado de un modelo de análisis discriminante.

Quinto paso: Interpretación de los resultados

Si la función discriminante es estadísticamente significativa y la precisión en la clasificación es aceptable, el investigador se debe centrar en realizar adecuadas interpretaciones de los resultados. Dentro de este proceso se examinan las funciones discriminantes para determinar la importancia relativa de cada variable independiente en la discriminación de los grupos. Se han propuesto tres métodos para determinar la importancia relativa: (1) las ponderaciones discriminantes estandarizadas, (2) las cargas discriminantes (correlaciones de estructura) y (3) los valores parciales de la F.

Ponderaciones discriminantes

El enfoque tradicional para interpretar las funciones discriminantes estudia el signo y la magnitud de la **ponderación discriminante** estandarizada (denominado algunas veces **coeficiente discriminante**) asignada a cada variable para calcular las funciones discriminantes. Cuando se ignora el signo, cada ponderación representa la contribución relativa de su variable asociada a esa función. Las variables independientes con ponderaciones relativamente grandes contribuyen más a la capacidad discriminante de la función que las variables con ponderaciones más pequeñas. El signo solamente denota que la variable ofrece una contribución positiva o negativa [3].

La interpretación de las ponderaciones discriminantes es análoga a la interpretación de las ponderaciones beta en el análisis de regresión y por ello está sujeta a las mismas críticas. Por ejemplo, una ponderación pequeña puede indicar, bien que su correspondiente variable es irrelevante para determinar una relación, o bien que ha sido apartada de la relación debido a un alto grado de multicolinealidad. Otro problema en la utilización de las ponderaciones discriminantes es que están sujetas a una considerable inestabilidad. Estos problemas implican que se tenga precaución en el uso de las ponderaciones para interpretar los resultados del análisis discriminante.

Cargas discriminantes

En los últimos años se ha incrementado la utilización de las cargas como fundamento de la interpretación debido a las deficiencias al utilizar ponderaciones. Las **cargas discriminantes**, denominadas también **correlaciones de estructura**, miden la correlación lineal simple entre cada variable independiente y la función discriminante. Las cargas discriminantes reflejan la varianza que las variables independientes comparten con la función discriminante, y pueden ser interpretadas como cargas de los factores para valorar la contribución relativa de cada variable independiente a la función discriminante. (En el Capítulo 3 se trata más profundamente la interpretación de la carga del factor.)

Las cargas discriminantes (al igual que las ponderaciones) pueden estar sujetas a inestabilidad. Se considera que las cargas son relativamente más válidas que las ponderaciones como medio de interpretación de la capacidad discriminante de las variables independientes debido a su naturaleza de correlación. El investigador debe ser cauto cuando utilice las cargas para interpretar las funciones discriminantes.

Valores parciales de la *F*

Como se discutió anteriormente, se pueden aplicar dos enfoques de cálculo —simultáneo y por etapas— para derivar las funciones discriminantes. Cuando se selecciona el método por etapas, se cuenta con un medio adicional de interpretar la capacidad discriminatoria de las variables independientes por medio del uso de los valores parciales de la F. Este se realiza examinando las tamaños absolutos de los valores significativos de la F y clasificándolos. Valores de la F grandes indican una capacidad discriminante mayor. En la práctica, las clasificaciones que emplean el enfoque de los valores de la F son las mismas que la clasificación derivada al utilizar las ponderaciones, pero los valores de la F indican además los niveles de significación asociados a cada variable.

Interpretación de dos o más funciones

Cuando existen dos o más funciones discriminantes significativas, nos enfrentamos con problemas adicionales de interpretación. Primero, ¿podemos simplificar las ponderaciones o las cargas discriminantes para facilitar la definición de cada función? Segundo, ¿cómo representamos la influencia de cada variable entre las funciones? Estos problemas se dan tanto en la medida de los efectos discriminantes totales entre funciones como al valorar el papel de cada variable en perfilar cada fun-

ción de forma separada. Tratamos estas dos cuestiones en los siguientes apartados introduciendo los conceptos de rotación de las funciones, índice de potencia y vectores de atributos extendidos en las representaciones gráficas.

Rotación de las funciones discriminantes

Después de haber construido las funciones discriminantes, éstas pueden ser «rotadas» para redistribuir la varianza. (El concepto está explicado de forma más detallada en el Capítulo 3.) Básicamente, la rotación mantiene la estructura original y la fiabilidad de la solución discriminante mientras que al mismo tiempo hace que las funciones sean más fáciles de interpretar de forma sustancial. En la mayoría de los casos, se hace uso de la rotación VARIMAX como fundamento de la rotación.

Índice de potencia

En las secciones anteriores, discutimos el uso de las ponderaciones estandarizadas y de las cargas discriminantes como medidas de la contribución de la variable a la función discriminante. Sin embargo, cuando se obtienen dos o más funciones, una medida resumen o compuesta es útil para describir las contribuciones de la variable entre todas las funciones significativas. El **índice de potencia** es una medida relativa entre todas las variables que señala la capacidad discriminante de cada variable [15]. Incluye tanto la contribución de la variable a la función discriminante (su carga discriminante) como la contribución relativa de la función a la solución gobal (una medida relativa entre los autovalores de las funciones). La composición es simplemente la suma de los índices de potencia individuales entre todas las funciones discriminantes significativas. Sin embargo, la interpretación de la medida compuesta esta limitada por el hecho de que es útil solamente para describir la posición relativa (como el orden en la clasificación) de cada variable, y el valor absoluto no tiene un verdadero significado. El índice de potencia se calcula a partir del siguiente proceso en dos etapas:

Etapa 1: Calcular un valor de potencia para cada función significativa. En el primer paso, la capacidad discriminante de la variable, representada por el valor al cuadrado de la carga discriminante, es «ponderada» por la contribución relativa de la función discriminante en la solución global. Primero, la medida del autovalor relativa a cada función significativa se calcula simplemente como,

$$\text{Autovalor relativo a la función discriminante } i = \frac{\text{Autovalor de la función discriminante } i}{\text{Suma de autovalores entre todas las funciones significativas}}$$

El valor de potencia de cada variable en una función discriminante es entonces,

$$\text{Valor de potencia de la variable } i \text{ para cada función } j = (\text{Carga discriminante}_{ij})^2 \times \text{Autovalor relativo de la función } j$$

Etapa 2: Calcular un índice de potencia compuesto entre todas las funciones significativas. Una vez que se ha calculado un valor de potencia para cada función, el índice de potencia compuesto se calcula como la suma de los valores de potencia de cada función discriminante significativa. El índice de potencia representa ahora el efecto discriminante total de la variable entre todas las funciones discriminantes significativas. Recuérdese que es solamente una medida relativa, y su valor absoluto no tiene un significado sustantivo.

Representación gráfica de las cargas discriminantes

Para reflejar las diferencias de los grupos en las variables predictoras, el investigador puede representar gráficamente las cargas discriminantes. El enfoque más sencillo es dibujar las verdaderas cargas rotadas y no rotadas en un gráfico. Lo ideal sería dibujar las cargas rotadas. Sin embargo, un enfoque incluso más preciso incluye lo que se denomina **extender los vectores**.

Antes de explicar el proceso de extensión, primero debemos definir un vector en este contexto. Un **vector** es meramente una línea recta dibujada desde el origen (centro) de un gráfico hasta las coordenadas de una determinada carga de una variable. La longitud de cada vector indica la importancia relativa de cada variable para discriminar entre los grupos. Para extender un vector, el investigador multiplica la carga discriminante (preferiblemente después de la rotación) por su respectivo valor univariante de la F.

El proceso de representación engloba a todas las variables incluidas en el modelo como significativas. Pero el investigador también puede representar las otras variables con ratios univariantes de la F significativos que no lo fueron en la función discriminante. Este proceso muestra la importancia de variables colineales que no están incluidas, al igual que en la solución por etapas. Empleando este procedimiento, notamos que los vectores señalan a los grupos que tienen las medias más altas sobre el respectivo predictor, y están lejos de los grupos que tienen las puntuaciones medias más bajas. Los centroides de grupo son también extendidas en este procedimiento multiplicándolas por el valor aproximado de la F asociado a cada función discriminante. Si las cargas son extendidas, los centroides deben ser también extendidas para representarlas exactamente en el mismo gráfico. Los valores de la F apropiados para cada función discriminante se obtienen a partir de la siguiente expresión:

$$\text{Valor de la}_{\text{función}_i} = \text{Autovalor}_{\text{función}_i}\left(\frac{\text{Tamaño muestral utilizado en la estimación} - \text{N.º de grupos}}{\text{N.º de grupos} - 1}\right)$$

Como ejemplo, suponga que la muestra de 50 observaciones se dividió en tres grupos. El múltiplo de cada autovalor sería $(50 - 3)/(3 - 1) = 23{,}5$. Para más detalles sobre este procedimiento, véase Dillon y Golstein [3].

Para quienes no desean extender los centroides y los vectores de atributos, se cuenta con los «mapas territoriales» que ofrecen la mayoría de los programas. Esta opción no incluye los vectores, pero sí representa los centroides y las cotas para cada grupo.

¿Cuál es el método de interpretación a emplear?

Se consideran varios métodos para interpretar la naturaleza de las funciones discriminantes, tanto para soluciones únicas como múltiples. ¿Qué métodos deberán emplearse? El enfoque de las cargas es algo más válido que el uso de las ponderaciones y debería ser utilizado siempre que fuera posible. El uso de valores de la F parciales y univariantes permite al investigador la utilización de varias medidas y la búsqueda de algo de consistencia en las evaluaciones de las variables. Si se estiman dos o más funciones, el investigador puede emplear varias herramientas gráficas y el índice de potencia, el cual ayuda a interpretar la solución multidimensional. Lo más importante es que el investigador emplee todos los métodos disponibles para llegar a la interpretación más precisa.

Sexto paso: Validación de los resultados

El último paso del análisis discriminante comprende la validación de los resultados discriminantes para asegurar que los resultados tienen validez tanto externa como interna. Dada la propensión del análisis discriminante a aumentar el ratio de aciertos si se evalúa solamente utilizando la muestra de análisis, la validación cruzada es una etapa fundamental. Muy a menudo la validación

cruzada se realiza con la muestra original, pero es posible emplear una muestra adicional como la ampliación de la muestra. Además de la validación cruzada, el investigador debe llevar a cabo el diseño de grupos que aseguren que las medias de estos grupos son indicadores válidos del modelo conceptual empleado para seleccionar las variables independientes. Estos dos enfoques se consideran a continuación.

División de la muestra o procedimientos de validación cruzada

Recuérdese que el procedimiento más frecuentemente utilizado para validar la función discriminante es dividir los grupos aleatoriamente en la muestra de análisis y en una ampliación de la muestra. Esto implica tener que construir una función discriminante con la muestra de análisis y después validarla con la ampliación de la muestra. La justificación para dividir la muestra total en dos grupos es que aparecerá un sesgo al alza en la capacidad predictiva de la función discriminante si los individuos incluidos en la construcción de la matriz de clasificación son los mismos que aquellos incluidos para calcular la función; es decir, la precisión clasificatoria será más alta que lo que es válido para la función discriminante, si fuese utilizada para clasificar una muestra separada. Las implicaciones de este sesgo al alza son particularmente importantes cuando el investigador está interesado en la validez externa de los resultados.

Sin embargo, otros investigadores han sugerido que se podría tener más confianza en la validez de la función siguiendo este proceso varias veces [14]. En lugar de dividir el total de la muestra aleatoriamente en muestra de análisis y ampliación de la muestra una sola vez, el investigador dividiría la muestra completa aleatoriamente en muestra de análisis y ampliación de la muestra varias veces, cada vez comprobando la validez de la función a través de la construcción de matrices de clasificación y de un ratio de aciertos. Entonces los diferentes ratios de aciertos serían promediados para obtener una medida única.

Se han sugerido métodos más sofisticados basados en la estimación con subconjuntos múltiples de la muestra para validar las funciones discriminantes [2, 3]. Los dos enfoques más ampliamente empleados son el método-*U* y el método *jackknife*. Ambos métodos están basados en el principio «dejar-uno-fuera», donde la función discriminante es ajustada con muestras tomadas repetidamente de la muestra original. El uso más extendido de este método ha sido estimar $k - 1$ muestras, eliminando una observación cada vez de una muestra de k casos. La diferencia principal entre los dos métodos es que el método-*U* se centra en la precisión clasificadora, mientras que el enfoque *jackknife* centra su atención en la estabilidad de los coeficientes discriminantes. Ambos métodos son bastante sensibles a tamaños muestrales pequeños. Se suele sugerir el uso de cualquiera de estos dos métodos solamente cuando el tamaño del grupo más pequeño sea al menos tres veces tan grande como el número de variables predictoras, y la mayoría de los investigadores proponen un ratio de cinco a uno [11]. A pesar de estas limitaciones, ambos métodos proporcionan la estimación más válida y consistente de la tasa de precisión clasificatoria. El uso de los métodos-*U* y *jackknife* ha sido limitado porque sólo uno de los paquetes informáticos principales los incluye.

Perfilar las diferencias entre los grupos

Otra técnica de validación consiste en perfilar los grupos de las variables independientes para asegurar su correspondencia con los fundamentos conceptuales empleados en la formulación del modelo original. Cuando el investigador ha identificado las variables independientes que contribuyen de forma más importante a la discriminación entre los grupos, el próximo paso es perfilar las características de los grupos atendiendo a las medias de los grupos. Este perfil permite al investigador comprender el carácter de cada grupo de acuerdo con las variables predictoras. Por ejemplo, refiriéndonos al conjunto de datos de HATCO presentado en la Tabla 5.1, vemos que la tasa media de «duración» para el grupo «compraría» es de 7,4, mientras que la tasa media comparable de «duración» para el grupo «no compraría» es de 3,2. Por ello un perfil de estos dos grupos mues-

tra que el grupo «compraría» clasifica la duración percibida del nuevo producto de forma sustancialmente más alta que el grupo «no compraría».

Otro enfoque es perfilar los grupos sobre un conjunto separado de variables que deben reflejar las diferencias de los grupos observados. Este perfil separador proporciona una valoración de validez externa en donde los grupos varían tanto en variable(s) independiente(s) como en el conjunto de variables asociadas. Esta es una característica parecida a la validación de los *clusters* estimados descrita en el Capítulo 9.

Regresión logística: Regresión con una variable dependiente binaria

Como se ha discutido, el análisis discriminante es apropiado cuando la variable dependiente es no métrica. Sin embargo, cuando la variable dependiente tiene sólo dos grupos, puede preferirse la regresión logística por varios motivos. En primer lugar, el análisis discriminante descansa sobre un cumplimiento estricto de los supuestos de normalidad multivariante y la igualdad de matrices de varianzas covarianzas entre los grupos, supuestos que no siempre se verifican. La regresión logística no se enfrenta a estos supuestos tan estrictos, y es mucho más robusta cuando estos supuestos no se cumplen, haciendo muy apropiada su aplicación en muchas situaciones. En segundo lugar, incluso si se cumplieran los supuestos, muchos investigadores prefieren la regresión logística porque es similar a la regresión. Ambas cuentan con contrastes estadísticos directos, capacidad para incorporar efectos no lineales y permitir un amplia variedad de diagnósticos. Por estas razones y otras más técnicas, la regresión logística es equivalente al análisis discriminante de dos grupos y puede considerarse más apropiada en muchas situaciones.

Nuestra presentación de la regresión logística no cubre cada una de las seis etapas del proceso de decisión; en su lugar resaltaremos las diferencias y parecidos entre la regresión logística y el análisis discriminante o la regresión múltiple. El análisis de la regresión múltiple se efectúa en el Capítulo 4.

Representación de la variable dependiente binaria

En el análisis discriminante, el carácter no métrico de una variable dependiente dicotómica se adecúa haciendo predicciones de pertenencia al grupo basadas en sus puntuaciones Z discriminantes. Esto requiere el cálculo de puntuaciones de corte y la asignación de observaciones a grupos. La regresión logística afronta esta tarea de forma algo parecida a la ya realizada en la regresión múltiple. Se diferencia de la regresión múltiple, sin embargo, en que predice directamente la probabilidad de ocurrencia de un suceso. Aunque el valor de la probabilidad sea una medida métrica, existen diferencias fundamentales entre la regresión múltiple y la logística. Los valores de la probabilidad pueden ser cualesquiera entre cero y uno, pero el valor predicho debe estar acotado para que caiga en el rango de cero y uno. Para definir una relación acotada por cero y uno, la regresión logística utiliza una relación supuesta entre las variables dependientes e independientes que recuerda a una curva en forma de S (véase Figura 5.9). Para niveles muy bajos de la variable independiente, la probabilidad se aproxima a cero. Según crece la variable independiente, la probabilidad crece a lo largo de la curva, pero como la pendiente empieza a decrecer para cierto nivel de la variable independiente, la probabilidad se acercará a uno sin llegar a excederlo. Como vimos en nuestra discusión de la regresión en el Capítulo 4, los modelos de regresión lineal no permitían captar tal relación, al ser inherentemente no lineal. Además, tales situaciones no pueden estudiarse mediante la regresión ordinaria, porque al hacerlo se incumplen varios supuestos. En primer lugar, el término de error de una variable discreta sigue una distribución binomial en lugar de la distribución normal, invalidando todos los contrastes estadísticos basados en el supuesto de normalidad. En segundo lugar, la varianza de una variable dicotómica no es constante, creando en consecuencia si-

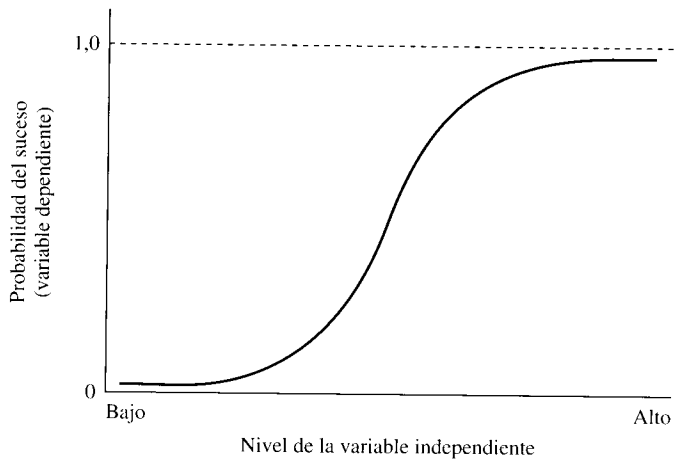


FIGURA 5.9. Forma de la relación logística entre las variables independiente y dependiente.

tuaciones de heterocedasticidad. La regresión logística se desarrolló para tratar precisamente estas situaciones. La única relación entre variables dependientes e independientes precisa de una aproximación algo diferente en la estimación, la evaluación de la bondad de ajuste y la interpretación de los coeficientes.

Estimación del modelo de regresión logística

La regresión logística, aunque incluya un único valor teórico resumen de los coeficientes estimados para cada variable independiente —como se encontró en la regresión múltiple— se estima de forma totalmente diferente. La regresión múltiple emplea el método de los mínimos cuadrados, que minimiza las sumas de las diferencias al cuadrado entre los valores reales y predichos de la variable dependiente. La naturaleza no lineal de la transformación logística requiere que otro procedimiento, el de máxima verosimilitud, se utilice de forma iterativa para encontrar la estimación «más probable» de los coeficientes. Por ello, se usa el **valor de la verosimilitud** en lugar de la suma de los cuadrados al calcular la medida de ajuste global del modelo.

El proceso de estimación de los coeficientes es, aun así, bastante parecido en muchos aspectos al de regresión lineal. Como se vio antes, el modelo logístico tiene la forma concreta de una **curva logística**. Para estimar un modelo de regresión logística, se ajusta esta curva a los datos reales. La Figura 5.10 representa dos ejemplos hipotéticos de ajuste de una relación logística a datos muestrales. Los datos reales, que un suceso tenga o no lugar (1 o 0), se representan como observaciones en lo alto o en lo bajo del gráfico. Estos son los sucesos que ocurren para cada valor de la variable independiente (el eje de las X). En la parte A, la curva logística no puede ajustar los datos bien porque hay varios valores de la variable independiente que cuentan tanto con sucesos como con no sucesos (esto es, un importante solapamiento de las distribuciones). Sin embargo, en la parte B, existe una relación mucho más definida, y la curva logística se ajusta a los datos bastante bien. Este sencillo ejemplo, similar a una nube de puntos entre las variables dependiente e independiente de la regresión con una línea que representa el «mejor ajuste» de la correlación, puede extenderse para incluir múltiples variables independientes como en la regresión.

Interpretación de los coeficientes

Una de las ventajas de la regresión logística es que sólo necesitamos saber si un suceso ocurrió (comprar o no, riesgo de crédito o no, quiebra de la empresa o éxito) para entonces utilizar un valor dicotómico como nuestra variable dependiente. A partir de este valor dicotómico, el procedimiento predice su estimación de la probabilidad de que el suceso tenga o no lugar. Si la predic-

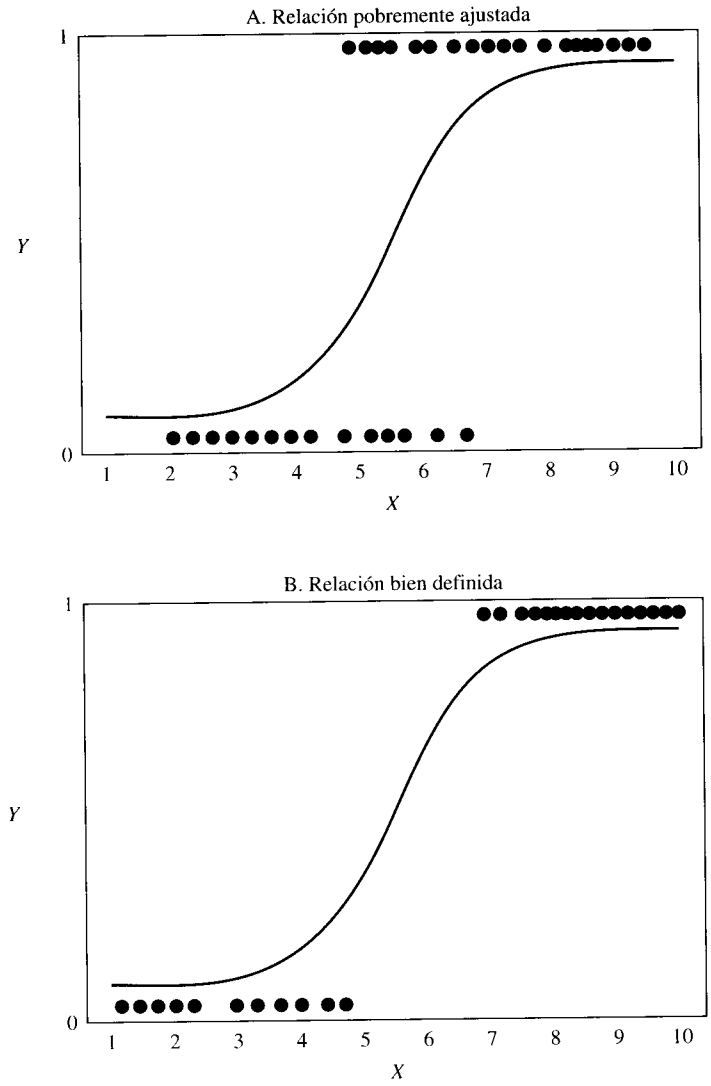


FIGURA 5.10. Ejemplos de ajuste de la curva logística a datos muestrales.

ción de la probabilidad es mayor que 0,50, entonces la predicción es sí, y no en otro caso. La regresión logística deriva su nombre de la **transformación logística** utilizada con la variable dependiente. Cuando se utiliza esta transformación, sin embargo, la regresión logística y sus coeficientes tienen un sentido diferente del que encontramos en la regresión con una variable dependiente métrica.

El procedimiento que calcula el **coeficiente logístico** compara la probabilidad de la ocurrencia de un suceso con la probabilidad de que no ocurra. Este **odds ratio** puede expresarse como:

$$\frac{\text{Prob}_{(\text{evento})}}{\text{Prob}_{(\text{no evento})}} = e^{B_0 + B_1X_1 + \cdots + B_nX_n}$$

Los coeficientes estimados (B_0, B_1, B_2, ..., B_n) son en realidad medidas de los cambios en el ratio de probabilidades, denominado *odds ratio*. Más aún, están expresados en logaritmos, por lo que necesitaríamos retransformarlos (tomando los valores del antilogaritmo) de tal forma que se evalúe más fácilmente su efecto sobre la probabilidad. Los programas de computador lo hacen automáticamente calculando tanto el coeficiente real como el transformado. Utilizar este procedimiento no cambia en modo alguno la forma de interpretar el signo del coeficiente. Un coeficiente

positivo aumenta la probabilidad, mientras que un valor negativo disminuye la probabilidad predicha.

Veamos un sencillo ejemplo para mostrar lo que estamos diciendo. Si B_i es positivo, su transformación (antilog) será mayor a 1, y el *odds ratio* aumentará. Este aumento se produce cuando la probabilidad prevista de ocurrencia de un suceso aumenta y la probabilidad prevista de su no ocurrencia disminuye. Por tanto, el modelo tiene una elevada probabilidad de ocurrencia. De la misma forma, si B_i es negativo, el antilogaritmo es menor que uno y el *odds ratio* disminuye. Un coeficiente cero equivale a un valor de 1,0, lo que no produce cambios en el *odds*. En varios textos [10] puede encontrarse una exposición más detallada de la interpretación de los coeficientes, la trasnformación logística y los procedimientos de estimación.

En nuestra anterior discusión sobre la distribución supuesta de las posibles variables dependientes, describimos una curva con forma de S o logística. Para representar esa relación entre las variables dependiente e independiente, los coeficientes deben representar efectivamente relaciones no lineales entre las variables dependientes e independientes. Aunque el proceso de transformación de tomar logaritmos proporciona una linealización de la relación, el investigador debe recordar que los coeficientes representan en realidad diferentes pendientes en la relación entre los valores de la variable independiente. De esta forma, puede estimarse la relación en forma de S. Si el investigador está interesado en la pendiente de la relación para varios niveles de la variable independiente, se pueden calcular los coeficientes y evaluar la relación [5].

Valoración de la bondad del ajuste del modelo estimado

La regresión logística es similar a la regresión múltiple en muchos otros resultados, pero es diferente en el método de estimación de los coeficientes. En lugar de minimizar la desviación de los cuadrados (mínimos cuadrados), la regresión logística maximiza la «verosimilitud» de que un suceso tenga lugar. La utilización de esta técnica de estimación alternativa requiere también que evaluemos el ajuste del modelo de varias formas.

La medida global de cómo se ajusta el modelo, similar al valor de la suma de errores o residuos al cuadrado en la regresión múltiple, viene dada por el valor de la verosimilitud. (Que es –2 veces el logaritmo del valor de verosimilitud y se representa por –2LL o –2 veces el logaritmo de la verosimilitud.) Un modelo con un buen ajuste tendrá un valor pequeño para –2LL. El valor mínimo para –2LL es cero. (Un ajuste perfecto tiene una verosimilitud de 1 y –2LL es cero.) El valor de la verosimilitud puede compararse asimismo entre ecuaciones, donde la diferencia representa el cambio en el ajuste predictivo desde una ecuación a otra. Los programas estadísticos cuentan con contrastes automáticos para evaluar la significación de estas diferencias.

El contraste chi-cuadrado para la reducción en el logaritmo del valor de verosimilitud proporciona una medida de mejora debida a la introducción de variable(s) independiente(s). Un modelo nulo, que es similar a calcular el total de la suma de los cuadrados utilizando sólo la media, proporciona el punto de partida para la comparación. Además de las contrastaciones estadísticas de los test chi-cuadrado, se han construido varias medidas diferentes tipo R^2 para representar el ajuste global del modelo, como lo hace el coeficiente de determinación de la regresión múltiple. El investigador puede construir un valor «pseudo R^2» para la regresión logística similar al valor R^2 del análisis de regresión [5]. El R^2 de un modelo logit (R^2_{logit}) se calcula como:

$$R^2_{\text{logit}} = \frac{-2\,\text{L}\,\text{L}_{\text{nulo}} - (-2\,\text{L}\,\text{L}_{\text{model}})}{-2\,\text{L}\,\text{L}_{\text{nulo}}}$$

Podemos evaluar el ajuste global de forma similar a la regresión múltiple, y podemos hacer uso de varios métodos que utilizan la característica no métrica de la variable dependiente. En primer lugar, podemos utilizar el método de las matrices de clasificación del análisis discriminante para evaluar la exactitud predictiva en términos de pertenencia al grupo. Todas las medidas relacionadas con la aleatoriedad utilizadas previamente son de aplicación también aquí. En segungo lugar, Hosmer y Lebeshow [10] han desarrollado otros contrastes de clasificación. Los casos se

dividen primero en 10 clases aproximadamente iguales. Luego, el número de sucesos reales y predichos se compara en cada clase con el estadístico chi-cuadrado. Este contraste proporciona una medida global de exactitud predicitiva que no se basa en el valor de verosimilitud, sino en la predicción real de la variable dependiente. El uso correcto de este contraste requiere un tamaño de muestra adecuado para asegurar que cada grupo cuenta al menos con cinco observaciones y nunca cae por debajo de uno. Además, el estadístico chi-cuadrado es sensible al tamaño muestral, permitiendo, por tanto, que esta medida encuentre diferencias estadísticamente muy pequeñas cuando el tamaño muestral crece. El investigador debería hacer uso de todas estas medidas de ajuste para valorar esta técnica, que cuenta con aspectos tanto de la regresión múltiple como del análisis discriminante.

Contrastación de la significación de los coeficientes

La regresión logística puede contrastar también la hipótesis de que un coeficiente sea distinto de cero (el cero significa que el *odds ratio* no cambia y que la probabilidad no se ve afectada), como se hizo en la regresión múltiple. En la regresión múltiple, el valor del t se utiliza para valorar la significatividad de cada coeficiente. La regresión logística utiliza un estadístico diferente, el **estadístico de Wald**. Este proporciona la significación estadística para cada coeficiente estimado de tal forma que se pueden contrastar hipótesis igual que en la regresión múltiple.

Otras semejanzas con la regresión múltiple A pesar del hecho de utilizar una medida dependiente binaria y de que el resultado sea la predicción de pertenencia al grupo, el formato de la regresión logística es bastante parecido al de la regresión múltiple. Al igual que en la regresión, los datos categóricos y nominales pueden incluirse como variables independientes por medio de su codificación como variables ficticias. También encontramos los procedimientos de selección de modelo al igual que en la regresión múltiple (por etapas hacia delante y hacia atrás). Finalmente, para examinar con mayor claridad los resultados, también se cuenta con muchas medidas de diagnóstico, como los residuos, los gráficos de los residuos y medidas de influencia.

El investigador que se enfrenta con una variable dicotómica no necesita recurrir a métodos diseñados que dan cabida a las limitaciones de la regresión múltiple ni se ve obligado a emplear el análisis discriminante, especialmente si no se verifican sus supuestos estadísticos. La regresión logística salva estos problemas y proporciona un método para tratar directamente con esta situación de la forma más eficiente posible.

Un ejemplo ilustrativo de dos grupos

Para ilustrar la aplicación de un análisis discriminante de dos grupos, utilizaremos variables tomadas de la base de datos de HATCO introducida en el Capítulo 1. Este ejemplo analiza cada uno de los seis pasos del proceso de construcción de un modelo, sobre un problema de investigación especialmente adecuado para el análisis discriminante múltiple.

Primer paso: Objetivos del análisis discriminante

Recuérdese que unas de las características más habituales obtenidas por HATCO en su trabajo era una variable categórica que indicaba cuál era el modelo de compra que utilizaba una empresa: análisis del valor total frente a especificación de la compra. Las empresas que emplean el análisis del valor total evalúan cada aspecto de la compra, incluyendo tanto el producto como los servicios que se van a comprar. Por otro lado, la especificación de la compra define todas las características de los productos y servicios que se desean, y el vendedor entonces realiza una oferta para cubrir esos requisitos. Ambos enfoques son adecuados en ciertas situaciones, pero el equipo de dirección de

HATCO espera que las empresas que utilicen estos enfoques harán hincapié en diferentes características de los oferentes en su decisión de selección. El objetivo es identificar las percepciones de HATCO que difieren significativamente entre empresas que emplean estos dos métodos de compra. La compañía entonces sería capaz de adaptar las presentaciones de las ventas y las ventajas ofrecidas a lo que mejor se adecúe a las percepciones del comprador. Para realizar esto, se seleccionó el análisis discriminante con el fin de identificar aquellas percepciones de HATCO que mejor distinguen empresas que emplean cada uno de los métodos de compra.

Segundo paso: Diseño de la investigación mediante el análisis discriminante

La etapa donde se diseña la investigación se centra en tres cuestiones clave: selección de las variables dependiente e independientes, valoración de si el tamaño muestral es el adecuado para el análisis planeado y división de la muestra con fines de validación.

Selección de las variables dependientes e independientes

Debido a que la variable dependiente, el método de compra que emplea una empresa, es una variable categórica de dos grupos, el análisis discriminante es la técnica adecuada. La encuesta también recogió percepciones de HATCO que ahora pueden utilizarse para diferenciar entre los dos grupos de empresas. El análisis discriminante emplea como variables independientes las primeras siete variables de la base de datos (de X_1 a X_7) para discriminar entre empresas que aplican cada uno de los métodos de compra (X_{11}).

Tamaño muestral

La muestra de 100 observaciones, cuando se divide en muestra de análisis y ampliación de la muestra (validación), cumple el tamaño mínimo sugerido (de 5 a 1) para la aplicación del análisis discriminante, ofreciendo un ratio de 9 a 1 de observaciones respecto a variables independientes (60 observaciones para 7 variables independientes potenciales). Aunque este ratio se incrementaría a 15 a 1 si la muestra no se dividiera, resulta más importante validar los resultados que incrementar el número de observaciones en la muestra de análisis. Además, ambos grupos exceden el tamaño mínimo de 20 observaciones por grupo. Finalmente, los dos grupos de empresas comprenden 60 y 40 observaciones, haciéndolos bastante comparables en tamaño, de tal forma que no influye ni en el proceso de estimación ni en el de clasificación.

División de la muestra

En la discusión previa se ha hecho hincapié en la necesidad de validar la función discriminante con una muestra dividida o ampliación de muestra. Cada vez que se emplea una ampliación de la muestra, el investigador debe asegurarse de que los tamaños muestrales resultantes son suficientes para apoyar el número de predictores incluidos en el análisis. La base de datos de HATCO tiene 100 observaciones, y se decidió que una ampliación de la muestra de 40 observaciones sería suficiente para la validación, y se dejarían 60 observaciones para la estimación de la función discriminante. Es importante asegurar la aleatoriedad en la selección de la ampliación de la muestra para que cualquier ordenación de las observaciones no afecte a los procesos de estimación y validación. Las fichas de control necesarias tanto para la selección de la ampliación de la muestra como para la realización del análisis discriminante de dos grupos están reflejadas en el Apéndice A al final del texto.

Tercer paso: Supuestos del análisis discriminante

Los principales supuestos que se presentan en el análisis discriminante son los relativos a la construcción del valor teórico o función discriminante (normalidad, linealidad y multicolinealidad) y a la estimación de la función discriminante (matrices de varianzas y covarianzas iguales). La inspección de las variables dependientes para comprobar la normalidad, la linealidad y la multicolinealidad se explica en el Capítulo 2. Para los objetivos de nuestro ejemplo, se considera que estos supuestos se cumplen aceptablemente.

En la mayoría de los programas estadísticos se ofrecen dos o más contrastes para el supuesto de igualdad de las matrices de covarianzas o de dispersión consideradas en el Capítulo 2. El contraste más común es el de la **M de Box** (para más detalles, véase Capítulo 2). En el ejemplo de dos grupos, la significación de las diferencias en las matrices de covarianzas entre los dos grupos es de 0,0320. Incluso aunque la significación fuese menor de 0,05 (recuérdese que en este contraste el investigador busca valores por encima del nivel de significación deseado), la sensibilidad del contraste con respecto a otros factores distintos a las diferencias en las covarianzas (por ejemplo, normalidad de las variables e incrementos del tamaño muestral) hacen que éste sea un nivel aceptable. Sin embargo, se utilizan las estimaciones de las varianzas separadas en lugar de estimaciones conjuntas en la fase de clasificación con fines ilustrativos. No es necesario llevar a cabo ninguna corrección adicional antes de que la estimación de la función discriminante pueda realizarse.

Cuarto paso: Estimación de la función discriminante y valoración del ajuste global

Comenzaremos nuestro estudio del análisis discriminante de dos grupos examinando la Tabla 5.5, en donde se muestran las medias de los grupos de cada una de las variables independientes, construida a partir de las 60 observaciones que forman la muestra de análisis. Además de perfilar los dos grupos, podemos identificar también las variables con las mayores diferencias en las medias

TABLA 5.5. Estadísticos descriptivos de los grupos y contrastes de igualdad para el análisis discriminante de dos grupos

	Medias de los grupos para las variables independientes[b]							
Variable dependiente[a]	X_1	X_2	X_3	X_4	X_5	X_6	X_7	*Tamaño muestral*
0: Especificación de compra	2,23	2,97	6,87	5,16	2,58	2,56	8,47	22
1: Análisis del valor total	4,26	2,08	8,57	5,44	3,18	2,83	6,01	38
Total	3,51	2,41	7,95	5,33	2,96	2,73	6,91	60
	Desviaciones estándar para las variables independientes[b]							
0: Especificación de compra	1,05	1,19	0,76	0,82	0,94	0,58	0,95	
1: Análisis del valor total	1,10	1,12	1,28	1,32	0,50	0,92	1,32	
Total	1,46	1,21	1,38	1,16	0,75	0,82	1,68	
	Contrastes de igualdad de las medias de grupo[c]							
Lambda de Wilks	0,542	0,873	0,645	0,986	0,846	0,973	0,499	
Ratio F univariante	48,992	8,453	31,881	0,822	10,576	1,620	58,176	
Nivel de significación	0,000	0,005	0,000	0,368	0,002	0,208	0,000	

[a] X_{11} = especificación de compra.
[b] X_1 = rapidez del envío; X_2 = nivel de precios; X_3 = flexibilidad de precios; X_4 = imagen del fabricante; X_5 = servicio gerneral; X_6 = imagen del personal de ventas; X_7 = calidad del producto.
[c] Lambda de Wilks (estadístico U) y ratio F univariante con 1 y 58 grados de libertad.

de los grupos (X_1, X_3 y X_7). La Tabla 5.5 también muestra la lambda de Wilks y el ANOVA univariante utilizados para valorar la significación entre las medias de las variables independientes para los dos grupos. Estos contrastes indican que cinco de las siete variables independientes muestran diferencias univariantes significativas entre los dos grupos. Sólo X_4 (imagen del fabricante) y X_6 (imagen de los vendedores) no son significativamente distintas. La finalidad del análisis discriminante es definir el conjunto de variables que mejor discriminan entre grupos. Para ello debemos estimar la función discriminante.

Estimación de la función discriminante

Dado que el objetivo de este análisis era determinar qué variables son las más eficientes para discriminar entre empresas que emplean los dos métodos de compra, se utilizó un procedimiento por etapas. Si el objetivo hubiera sido simplemente determinar las capacidades discriminantes del conjunto completo de servicios, sin considerar la influencia de ninguno individualmente, todas las variables deberían haberse incluido en el modelo simultáneamente. Se hará uso de la medida D^2 de Mahalanobis en el procedimiento por etapas para determinar la variable con mayor capacidad discriminatoria.

El procedimiento por etapas comienza con todas las variables excluidas del modelo para seleccionar posteriormente aquella variable que maximiza la distancia de Mahalanobis entre los grupos. En este ejemplo, se requiere un valor de significación mínimo de 0,05 para ser incorporada, y se utiliza la D^2 de Mahalanobis para seleccionar realmente las variables. La distancia de D^2 de Mahalanobis máxima está asociada a X_7 (véase Tabla 5.6). Después de que X_7 se incorporase al mo-

TABLA 5.6. Resultados del paso 1 del modelo de análisis discriminante de dos grupos por pasos

PRIMER PASO: X_7 (CALIDAD DEL PRODUCTO) INCLUIDA EN EL ANÁLISIS

ESTADÍSTICOS RESUMEN

		Grados de libertad		*Significación*	*Entre los grupos*
Lambda de Wilks	0,499	1	1 58		
F equivalente	58,176		1 58	0,000	
Mínima D^2	4,175				0 y 1
F equivalente	58,176		1 58	0,000	

VARIABLES EN EL ANÁLISIS DESPUÉS DEL PRIMER PASO

		F para eliminar	
Variables	*Tolerancia*	*Valor*	*Significación*
X_7 calidad del producto	1,00	58,176	0,000

VARIABLES FUERA DEL ANÁLISIS DESPUÉS DEL PRIMER PASO

			F para entrar			
Variables	*Tolerancia*	*Tolerancia mínima*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_1 rapidez del envío	0,973	0,973	16,680	0,000	6,615	0 y 1
X_2 nivel de precios	0,933	0,933	0,454	0,503	4,242	0 y 1
X_3 flexibilidad de precios	0,997	0,997	18,196	0,000	6,837	0 y 1
X_4 imagen del fabricante	0,963	0,963	2,874	0,095	4,596	0 y 1
X_5 servicio global	0,994	0,994	7,203	0,010	5,229	0 y 1
X_6 imagen del personal de ventas	0,962	0,962	3,896	0,053	4,745	0 y 1

CONTRASTACIÓN DE LA SIGNIFICACIÓN DE LAS DIFERENCIAS DE GRUPO DESPUÉS DEL PRIMER PASO[a]

	Grupo 0: Especificación de la compra
Grupo 1: Análisis del valor total	58,176 (0,000)

[a] El estadístico *F* y el nivel de significación (en paréntesis) entre grupos después del primer paso. Cada estadístico *F* tiene 1 y 58 grados de libertad.

delo, las restantes variables se evaluaron en base a la distancia entre sus medias después de haber eliminado la varianza asociada con X_7. Una vez más, las variables con valores de significación mayores a 0,05 se excluyeron de ser consideradas en la siguiente fase.

Tres variables alcanzaron el criterio del nivel de significación del 0,05 para ser consideradas en la siguiente fase (X_1, X_3 y X_5). La variable X_3 es la siguiente mejor candidata para entrar en el modelo por tener la mayor D^2 de Mahalanobis (6,837) (véase Tabla 5.6). Dado el mayor valor de la D^2 de Mahalanobis de X_1 (6,615), es muy probable que entre en el modelo en una fase posterior si no está altamente correlacionada con variables previamente seleccionadas. Los contrastes de significación estadística deben calcularse después de que el efecto de la(s) variable(s) en los modelos se haya eliminado. Por ejemplo, una alta multicolinealidad de X_1 respecto a otras variables del modelo podría reducir sustancialmente el nivel de significación y la D^2 de Mahalanobis. También, en casos donde se introducen dos o más variables en el modelo, las variables que ya están en el modelo se evalúan para su posible eliminación. Se puede eliminar una variable si existe una alta multicolinealidad entre ella y otras variables independientes ya incluidas tal que su nivel de significación caiga por debajo de su nivel de significación para ser excluida (0,10).

En la segunda etapa (véase Tabla 5.7) X_3 entra en el modelo como se esperaba. Al igual que en la primera etapa, el modelo completo es significativo ($F = 46,81$), como también lo es la capa-

TABLA 5.7. Resultados del paso 2 del modelo de análisis discriminante de dos grupos por pasos

SEGUNDO PASO: X_3 (FLEXIBILIDAD DE PRECIOS) INCLUIDA EN EL ANÁLISIS

ESTADÍSTICOS RESUMEN

		Grados de libertad		*Significación*	*Entre los grupos*
Lambda de Wilks	0,378	2	1 58		
F equivalente	46,810		2 57	0,000	
Mínima D^2	6,837				0 y 1
F equivalente	46,810		2 57	0,000	

VARIABLES EN EL ANÁLISIS DESPUÉS DEL SEGUNDO PASO

		F para eliminar			
Variables	*Tolerancia*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_3 flexibilidad de precios	0,997	18,196	0,000	2,288	0 y 1
X_7 calidad del producto	0,997	40,195	0,000	4,175	0 y 1

VARIABLES FUERA DEL ANÁLISIS DESPUÉS DEL SEGUNDO PASO

			F para entrar			
Variables	*Tolerancia*	*Tolerancia mínima*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_1 rapidez del envío	0,932	0,932	7,974	0,007	8,403	0 y 1
X_2 nivel de precios	0,809	0,809	0,661	0,419	6,967	0 y 1
X_4 imagen del fabricante	0,946	0,946	3,884	0,054	7,600	0 y 1
X_5 servicio global	0,980	0,980	7,770	0,007	8,363	0 y 1
X_6 imagen del personal de ventas	0,959	0,958	3,557	0,064	7,536	0 y 1

CONTRASTACIÓN DE LA SIGNIFICACIÓN DE LAS DIFERENCIAS DE GRUPO DESPUÉS DEL SEGUNDO PASO[a]

	Grupo 0: Especificación de la compra
Grupo 1: Análisis del valor total	46,810 (0,000)

[a] El estadístico F y el nivel de significación (en paréntesis) entre grupos después del segundo paso. Cada estadístico F tiene 2 y 57 grados de libertad.

cidad discriminante de las variables ahora incluidas (X_3 y X_7). Como se indicó anteriormente, X_1 es la próxima candidata para ser incluida, pero el valor de la F para entrar se ha reducido de forma importante debido a la multicolinealidad de X_1 con X_3 y X_7. Además la D^2 de Mahalanobos se ha incrementado (de 4,175 a 6,837) indicativa de un «reparto» y separación de los grupos debida a X_3 y X_7 ya incluidas en la función discriminante. Nótese que X_5 es casi idéntica en la capacidad discriminante restante, pero X_1 entrará en la tercera etapa debido a su ligera ventaja. La Tabla 5.8 revisa los resultados de la tercera fase del proceso por etapas, donde X_1 no entra en la función discriminante. Los resultados globales son todavía estadísticamente significativos y continúan mejorando la discriminación, como lo prueba el valor de la lambda de Wilks (de 0,378 a 0,331). Incluidas X_1, X_3 y X_7, ninguna de las cuatro variables independientes restantes pasa el criterio de entrada de 0,05. Después de incluir X_1 en la ecuación, X_6 añade una capacidad discriminatoria relativamente escasa y no cumple el criterio de entrada. Por tanto, el proceso de discriminación finaliza con tres variables (X_1, X_3 y X_7) constituyendo la función discriminante.

La Tabla 5.9 refleja los resultados globales del análisis discriminante por etapas después de que todos los discriminantes significativos se han incluido en la estimación de la función discriminante. Esta Tabla resumen describe tres variables (X_1, X_3 y X_7) que son discriminadores significativos basados en su lambda de Wilks y los valores mínimos de la D^2 de Mahalanobis. Los aspectos mul-

TABLA 5.8. Resultados del paso 2 del modelo de análisis discriminante de dos grupos por pasos

TERCER PASO: X_1 (RAPIDEZ EN EL ENVÍO) INCLUIDA EN EL ANÁLISIS

ESTADÍSTICOS RESUMEN

		Grados de libertad			*Significación*	*Entre los grupos*
Lambda de Wilks	0,331	3	1	58		
F equivalente	37,683		3	56	0,000	
Mínima D^2	8,403					0 y 1
F equivalente	37,683		3	56	0,000	

VARIABLES EN EL ANÁLISIS DESPUÉS DEL TERCER PASO

		F para eliminar			
Variables	*Tolerancia*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_7 calidad del producto	0,965	.	0,000	4,886	0 y 1
X_3 flexibilidad de precios	0,954	.	0,004	6,615	0 y 1
X_1 rapidez del envío	0,932	.	0,007	6,387	0 y 1

VARIABLES FUERA DEL ANÁLISIS DESPUÉS DEL TERCER PASO

			F para entrar			
Variables	*Tolerancia*	*Tolerancia mínima*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_2 nivel de precios	0,788	0,788	.	0,238	8,728	0 y 1
X_4 imagen del fabricante	0,937	0,920	.	0,120	8,972	0 y 1
X_5 servicio global	0,570	0,542	.	0,248	8,716	0 y 1
X_6 imagen del personal de ventas	0,957	0,925	.	0,109	9,010	0 y 1

CONTRASTACIÓN DE LA SIGNIFICACIÓN DE LAS DIFERENCIAS DE GRUPO DESPUÉS DEL TERCER PASO[a]

	Grupo 0: Especificación de la compra
Grupo 1: Análisis del valor total	37,683 (0,000)

[a] El estadístico F y el nivel de significación (en paréntesis) entre grupos después del tercer paso. Cada estadístico F tiene 2 y 57 grados de libertad.

TABLE 5.9. Resumen de los resultados del análisis discriminante de dos grupos

TABLA RESUMEN

	Acción		*Lambda de Wilks*		*Mínima D^2*		
Etapas	*Entró*	*Se eliminó*	*Valor*	*Significación*	*Valor*	*Significación*	*Entre los grupos*
1	X_7 calidad del producto		0,499	0,000	4,175	0,000	0 y 1
2	X_3 flexibilidad de precios		0,378	0,000	6,837	0,000	0 y 1
3	X_1 rapidez del envío		0,331	0,000	8,403	0,000	0 y 1

FUNCIONES DISCRIMINANTES CANÓNICAS

		Porcentaje de varianza							
Función	*Autovalor*	*Función*	*Acumulado*	*Correlación canónica*	*Tras la función*	*Lambda de Wilks*	*Chi-cuadrado*	*df*	*Significación*
					0	0,331	62,424	3	0,000
1[a]	2,019	100	100	0,818					

COEFICIENTES DE LAS FUNCIONES DISCRIMINANTES CANÓNICAS

Variables independientes	*Estandarizada*	*No estandarizada*
X_1 rapidez en el envío	0,447	0,413
X_3 flexibilidad de precios	0,472	0,421
X_7 calidad del producto	−0,659	−0,549
Constante		−1,003

MATRIZ DE ESTRUCTURA[b]

Variables independientes	*Cargas de la función: Función* 1
X_7 calidad del producto	−0,705
X_1 rapidez en el envío	0,647
X_3 flexibilidad de precios	0,522
X_2 nivel de precios	−0,443
X_6 imagen del personal de ventas	−0,168
X_4 imagen del fabricante	0,155
X_5 servicio global	−0,145

COEFICIENTES DE LA FUNCIÓN DE CLASIFICACIÓN

Variables independientes	*Grupo* 0: *Especificación de compra*	*Grupo* 1: *Análisis del valor total*
X_1 rapidez en el envío	2,021	3,219
X_3 flexibilidad de precios	4,728	5,950
X_7 calidad del producto	5,932	4,340
Constante	−44,606	−45,848

MEDIA DE LOS GRUPOS (CENTROIDES) DE LAS FUNCIONES DISCRIMINANTES CANÓNICAS

Grupo	*Centroides de grupo: Función* 1
Especificación de compra	−1,836
Valor total de compra	1,063

[a] Señala la función discriminante canónica 1 que queda en el análisis.
[b] Correlaciones entre grupos entre variables discriminantes y funciones discriminates canónicas (variables ordenadas por tamaño de la correlación entre la función).

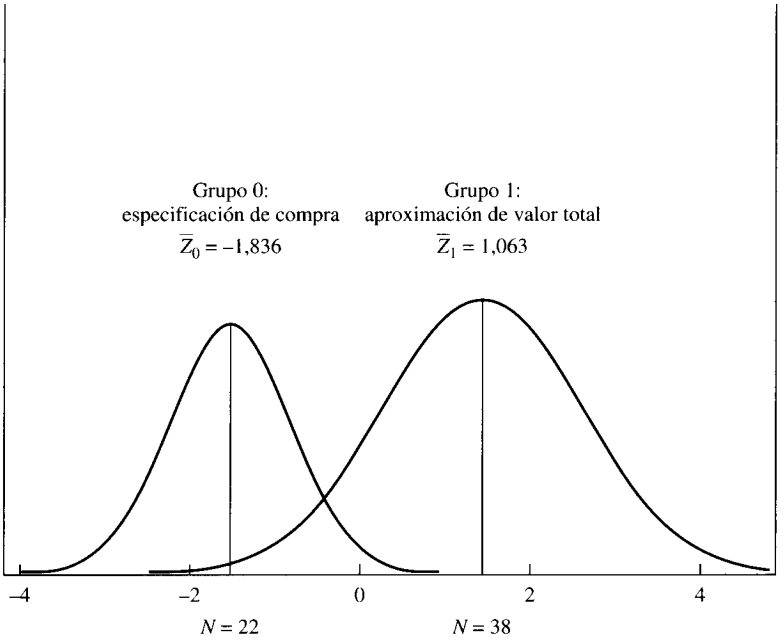


FIGURA 5.11. Gráfico de las centroides de grupo (*Z*).

tivariantes del modelo están reflejados bajo el título «Funciones discriminantes canónicas». Nótese que la función discriminante es altamente significativa (0,000) y presenta una correlación canónica de 0,818. Esta correlación se interpreta elevándola al cuadrado $(0{,}818)^2 = 0{,}669$, y concluyendo que el 66,9 por ciento de la varianza de la variable dependiente (X_{11}) puede ser explicado por este modelo, el cual incluye solamente tres variables independientes. Se ofrecen también los coeficientes de la función discriminante canónica estandarizada, pero resultan menos preferidos en la fase de interpretación que las cargas discriminantes. Los coeficientes discriminantes no estandarizados se utilizarán para calcular las puntuaciones *Z* discriminantes que pueden usarse en la clasificación. Las cargas discriminantes se presentan bajo el título «Matriz de estructura» y están ordenados de mayor a menor según el tamaño de la carga. Las cargas serán comentadas más adelante en la etapa de interpretación. Los coeficientes de la función de clasificación, también conocidas como las funciones discriminantes lineales de Fisher, se emplean al clasificar y se discuten más adelante. Finalmente, se presentan los centroides de los grupos, que representan las puntuaciones de la media de las puntuaciones de las funciones discriminantes individuales para cada grupo.

Los centroides de los grupos se pueden emplear para interpretar los resultados de la función discriminante desde una perspectiva general. La Tabla 5.9 revela que la centroide del grupo para las empresas que emplean la especificación de la compra (grupo 0) es –1.836, mientras la centroide del grupo de las empresas que emplean el enfoque del análisis del valor total (grupo 1) es 1,1063. La Figura 5.11 es una representación de los centroides que muestra la desviación de cada grupo respecto de la media global de los dos grupos. Para mostrar que la media total es 0, hay que multiplicar el número de empresas en cada grupo por su centroide y sumar el resultado (por ejemplo, $-1{,}836 \times 22 + 1{,}1063 \times 38 = 0{,}0$).

Valoración del ajuste global

El segundo paso en la etapa de estimación es valorar la capacidad predictiva de la función discriminante. Para realizar esto, debemos construir una matriz de clasificación. Se calculan las matrices de clasificación tanto para la muestra de análisis como para la ampliación de la muestra. Aunque el examen sobre la ampliación de la muestra y su capacidad predictiva se lleva a cabo

realmente en la etapa de validación, los resultados se discutirán ahora para facilitar la comparación entre la muestra de estimación y la ampliación de la muestra. Para comprender mejor el proceso de clasificación, debemos determinar la puntuación de corte, el criterio frente al cual se compara cada puntuación Z discriminante para cada observación para determinar en qué grupo debe ser clasificada.

En esta muestra de análisis de 60 observaciones, sabemos que la variable dependiente está formada por dos grupos, 22 empresas siguen el enfoque de especificación de la compra, y las restantes 38 empresas el método de análisis del valor total. Si no estuviéramos seguros de si las proporciones poblacionales están representadas en la muestra, entonces deberíamos emplear probabilidades iguales. Sin embargo, dado que nuestra muestra de empresas fue tomada aleatoriamente, podemos estar razonablemente seguros de que esta muestra refleja las proporciones poblacionales. Por ello, este análisis discriminante empleará las proporciones muestrales para especificar primeras probabilidades con la finalidad de clasificar.

Para ilustrar la importancia de la determinación de la puntuación de corte, centrémonos en cómo se utilizan las primeras probabilidades en el calculo de la puntuación de corte. Si los dos grupos son de igual tamaño, la puntuación de corte simplemente sería la media de los dos centroides. Debido a que los grupos son de diferentes tamaños, se debe emplear una media ponderada. La media ponderada se calcula como sigue:

$$Z_{CU} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

donde

Z_{CU} = puntuación de corte crítica para grupos de distintos tamaños
N_A = tamaño muestral del grupo A
N_B = tamaño muestral del grupo B
Z_A = centroide del grupo A
Z_B = centroide del grupo B

Sustituyendo los valores adecuados en la expresión, podemos obtener la puntuación de corte crítica (suponiendo igualdad de costes en caso de clasificación incorrecta):

$$Z_{CU} = \frac{38(-1{,}836) + 22(1{,}063)}{38 + 22} = -0{,}773$$

Los tamaños de los grupos empleados en el cálculo anterior están basados en el conjunto de datos empleando la muestra de análisis y no incluyen la ampliación de la muestra. El procedimiento de clasificación de las empresas mediante la puntuación de corte óptima es el siguiente:

1. Clasificar una empresa en el grupo de especificación de la compra si la puntuación discriminante es menor a $-0{,}773$.
2. Clasificar una empresa en el grupo del enfoque del análisis del valor total si su puntuación discriminante es mayor que $-0{,}773$. Se calcularon las matrices de clasificación para las observaciones tanto de la muestra de análisis como de la ampliación de la muestra. Los resultados aparecen en la Tabla 5.10. Como se anticipó, la muestra de análisis cuenta con un 91,7 por ciento de precisión predictiva, ligeramente superior al 86,4 por ciento de precisión de la ampliación de la muestra. A continuación examinamos las observaciones individuales y la precisión en su clasificación.

El 91,7 por ciento de precisión clasificatoria es bastante alto. A efectos ilustrativos, sin embargo, la compararemos con las otras dos medidas aleatorias de clasificación de los individuos de forma correcta sin usar la función discriminante. La primera medida es el criterio de aleatoriedad proporcional. Tenemos tamaños muestrales distintos y queremos identificar correctamente a los individuos de ambos grupos. El primer grupo, aquellos que usan especificación de la compra, constituye

TABLA 5.10. Matrices de clasificación del análisis discriminante de dos grupos para las muestras de análisis y ampliada

RESULTADOS DE LA CLASIFICACIÓN: MUESTRA DE ANÁLISIS[a]

		Pertenencia al grupo predicha	
Grupo real	*Número de casos*	*Especificación de compra*	*Análisis del valor total*
Especificación de compra	22	21	1
		95,5%	4,5%
Análisis del valor total	38	4	34
		10,5%	89,5%
Número de casos		25	35
Análisis del valor total	38	4	34

RESULTADOS DE LA CLASIFICACIÓN: AMPLIACIÓN DE LA MUESTRA[b]

		Pertenencia al grupo predicha	
Grupo real	*Número de casos*	*Especificación de compra*	*Análisis del valor total*
Especificación de compra	18	15	3
		83,3%	16,7%
Análisis del valor total	22	3	19
		13,6%	86,4%
Número de casos		18	22

[a] Porcentaje de los casos agrupados correctamente clasificados: 91,7 por ciento [(21 + 34)/60 = 0,917].
[b] Porcentaje de los casos agrupados correctamente clasificados: 86,4 por ciento [(15 + 19)/40 = 0,864].

el 36,7 por ciento de la muestra de análisis (22/60) con el segundo grupo, que usa el análisis del valor total, forma el restante 63,3 por ciento (38/60). La fórmula para el criterio de aleatoriedad proporcional es:

$$C_{PRO} = p^2 + (1 - p)^2$$

donde

C_{PRO} = el criterio de aleatoriedad proporcional
p = proporción de las empresas del grupo 1
$1 - p$ = proporción de las empresas del grupo 2

El valor de la aleatoriedad proporcional calculado es 0,535 ($0{,}367^2 + 0{,}633^2 = 0{,}535$).

El criterio de aleatoriedad máxima es simplemente el porcentaje correctamente clasificado si todas las observaciones fuesen clasificadas en el grupo con la probabilidad más elevada de ocurrencia. Debido a que el grupo 1 (análisis del valor total) ocurre el 63,3 por ciento de las veces, si asignamos todas las observaciones a este grupo, clasificaríamos de forma correcta el 63,3 por ciento. Dado que el criterio de aleatoriedad máxima es mayor que el criterio de contraste proporcional, nuestro modelo debería alcanzar el nivel del 63,3 por ciento.

La precisión clasificatoria del 91,7 por ciento es sustancialmente más alta que el 53,6 del criterio de aleatoriedad proporcional y que el 63,3 por ciento del criterio de máxima aleatoriedad. También supera el umbral recomendado del valor más el 25 por ciento que en este caso queda en el 79,1 por ciento (63,3 + 1,25 = 79,1). Nótese que el criterio de aleatoriedad proporcional se compara con el porcentaje correctamente clasificado en la ampliación de la muestra, reduciendo el sesgo por exceso que se observó en la clasificación de la muestra de análisis.

La medida final de la precisión en la clasificación es la Q de Press. De lo discutido anteriormente, el cálculo para la muestra de análisis es

$$Q \text{ de Press}_{\text{muestra de análisis}} = \frac{[60 - (56 \times 2)]^2}{60(2 - 1)} = 45{,}1$$

Y el cálculo para la ampliación de la muestra es

$$Q \text{ de Press}_{\text{ampliación de la muestra}} = \frac{[40 - (37 \times 2)]^2}{40(2 - 1)} = 28{,}9$$

En ambos casos, el valor calculado excede al valor crítico de 6,63. Por ello, la precisión clasificatoria para la muestra de análisis y, más importante, para la ampliación de la muestra exceden en un nivel de significatividad estadística la precisión clasificatoria esperada de forma aleatoria. El investigador debe usar siempre con cautela la ampliación de la muestra con conjuntos de datos pequeños. En este caso, el tamaño muestral de 40 fue adecuado, pero siempre son más deseables tamaños muestrales más grandes.

Diagnóstico por casos

Además de examinar los resultados globales, podemos examinar las observaciones individuales por su precisión predictiva e identificar en concreto las observaciones mal clasificadas. La Tabla 5.11 contiene las predicciones de grupo para las muestras de análisis y ampliación de la muestra. Los casos 88, 84, 93 y 82 están realmente en el grupo 1 (análisis del valor total) pero se ha predicho que estén en el grupo 0 (especificación de la compra). Estas cuatro clasificaciones erróneas se muestran en la Tabla 5.10. Además, el caso 13 está realmente en el grupo 0, pero se ha predicho que sea del grupo1. Éste también se muestra en la Tabla 5.10 como clasificación errónea. Puede llevarse a cabo un examen similar para la ampliación de la muestra.

Una vez que se identifican los casos mal clasificados, se puede llevar a cabo un análisis más extenso para comprender las razones de su clasificación errónea. En la Tabla 5.12 se combinan los casos de clasificación errónea de las muestras de análisis y ampliación de la muestra y se comparan luego con los casos correctamente clasificados. El objetivo está en identificar diferencias concretas en las variables independientes que puedan identificar, bien nuevas variables que añadir, o características comunes que puedan considerarse. Los cuatro casos (para ambas muestras) mal clasificados en el grupo de especificación de la compra cuentan con diferencias significativas sobre dos de las tres variables independientes en la función discriminante. Sobre X_7 y X_3, su perfil está más cercano a los miembros del grupo de análisis del valor total. También, los casos mal clasificados cuentan con una peor imagen de los vendedores (X_6). Esto puede sugerir que posiblemente fueran mal clasificados inicialmente o que otra variable pueda identificar otra característica que justifique su uso de especificación de la compra, incluso aunque parezca similar al grupo de análisis del valor total. El grupo de análisis del valor total muestra un patrón similar, aunque con variables diferentes. Es este caso, los siete casos mal clasificados están más próximos al grupo de especificación de la compra para dos variables (X_1 y X_3), pero no para X_7. Este grupo muestra también diferencias significativas importantes sobre tres de las otras cuatro variables independientes. El investigador debería examinar los patrones en ambos grupos con el objetivo de comprender las caracteríticas comunes a ellos para definir los motivos de la clasificación errónea.

Quinto paso: Interpretación de los resultados

Después de estimar la función, el siguiente paso es la interpretación. Esta etapa incluye el examen de la función para determinar la importancia relativa de cada variable independiente para discriminar entre los grupos. La Tabla 5.13 contiene, entre las medidas interpretativas, las ponderacio-

TABLA 5.11. Predicciones de grupo para los casos individuales en el análisis discriminante de dos grupos

Caso ID	Grupo real	Puntuación Z discriminante	Grupo predicho	Caso ID	Grupo real	Puntuación Z discriminante	Grupo predicho
MUESTRA DE ANÁLISIS							
79	0	−3,04	0	14	1	0,47	1
65	0	−3,01	0	11	1	0,51	1
39	0	−2,98	0	67	1	0,60	1
96	0	−2,67	0	90	1	0,63	1
48	0	−2,24	0	7	1	0,73	1
2	0	−2,22	0	95	1	0,73	1
6	0	−2,22	0	1	1	0,74	1
45	0	−2,11	0	80	1	0,79	1
54	0	−2,08	0	92	1	1,00	1
86	0	−1,96	0	58	1	1,04	1
71	0	−1,64	0	51	1	1,12	1
8	0	−1,64	0	15	1	1,37	1
31	0	−1,63	0	20	1	1,37	1
70	0	−1,55	0	49	1	1,42	1
68	0	−1,55	0	97	1	1,49	1
99	0	−1,51	0	43	1	1,53	1
52	0	−1,42	0	28	1	1,55	1
53	0	−1,32	0	59	1	1,56	1
36	0	−1,27	0	26	1	1,64	1
24	0	−1,14	0	29	1	1,65	1
89	0	−1,12	0	50	1	1,74	1
88	1	−0,97	0	72	1	1,85	1
84	1	−0,92	0	73	1	1,96	1
93	1	−0,81	0	81	1	2,24	1
82	1	−0,74	0	61	1	2,36	1
17	1	−0,69	1	47	1	2,40	1
32	1	−0,60	1	33	1	2,54	1
23	1	−0,54	1	25	1	2,68	1
12	1	−0,12	1	42	1	3,01	1
13	0	−0,06	1	5	1	3,05	1
AMPLIACIÓN DE LA MUESTRA							
94	0	−2,62	0	85	0	−0,08	1
98	0	−2,60	0	76	1	0,02	1
83	0	−2,28	0	38	1	0,47	1
37	0	−2,18	0	21	1	0,52	1
40	0	−2,15	0	100	1	0,53	1
3	0	−1,70	0	63	1	0,54	1
34	0	−1,52	0	66	1	0,58	1
75	0	−1,49	0	74	1	0,97	1
60	0	−1,46	0	18	1	1,02	1
30	0	−1,44	0	22	1	1,04	1
10	0	−1,39	0	9	1	1,05	1
41	0	−1,25	0	77	1	1,15	1
27	0	−1,19	0	55	1	1,16	1
4	0	−1,18	0	19	1	1,54	1
57	0	−1,13	0	46	1	1,56	1
91	1	−0,82	0	78	1	1,59	1
56	1	−0,79	0	16	1	1,85	1
64	1	−0,74	0	69	1	2,13	1
35	0	−0,17	1	62	1	2,51	1
87	0	−0,15	1	44	1	2,65	1

TABLA 5.12. Perfil de las observaciones correcta e incorrectamente clasificadas en el análisis discriminante de dos grupos

	Puntuaciones medias			*Contraste t*	
Variable dependiente Grupo/variables de perfil	*Correctamente clasificado*	*Mal clasificado*	*Diferencia*	*Valor de t*	*Significación*
Especificación de compra ($X_{11} = 0$)	($N = 36$)	($N = 4$)			
X_1 rapidez del envío[a]	2,494	2,550	−0,056	−0,274[b]	0,786
X_2 nivel de precios	3,089	2,075	1,014	1,681	0,101
X_3 flexibilidad de precios[a]	6,650	8,175	−1,525	−3,755	0,001
X_4 imagen del fabricante	5,358	4,775	0,583	0,838[b]	0,460
X_5 servicio global	2,758	2,325	0,433	0,895	0,376
X_6 imagen del personal de ventas	2,694	2,000	,694	2,279	0,028
X_7 calidad del producto[a]	8,483	6,575	1,908	4,918	0,000
Análisis del valor total ($X_{11} = 1$)	($N = 53$)	($N = 7$)			
X_1 rapidez del envío[a]	4,372	2,829	1,543	7,075[b]	0,000
X_2 nivel de precios	1,730	3,600	−1,870	−5,563	0,000
X_3 flexibilidad de precios[a]	8,881	6,657	2,224	4,663[b]	0,002
X_4 imagen del fabricante	5,057	6,400	−1,343	−2,331	0,052
X_5 servicio global	3,026	3,229	−0,203	−0,852	0,398
X_6 imagen del personal de ventas	2,609	3,314	−0,705	−2,079	0,042
X_7 calidad del producto[a]	5,981	6,914	−0,933	−1,830	0,072

[a] Variables incluidas en la función discriminante.
[b] Contraste t calculado con estimaciones de la varianza separadas, en lugar de con estimaciones de las varianzas agrupadas, (*pooled*) debido a que el contraste de Levene detectó diferencias significativas de las varianzas entre los grupos.

nes discriminantes, las cargas de la función y el ratio F univariante. Las variables independientes fueron seleccionadas con el procedimiento por etapas, y tres (X_1, X_2 y X_7) fueron suficientemente significativas como para incluirlas en la función. Con fines de interpretación, clasificamos las variables independientes en términos tanto de las ponderaciones como de las cargas —indicadores de su capacidad discriminante. Los signos no afectan a las clasificaciones; solamente indican una relación positiva o negativa con la variable dependiente. Como las cargas se consideran más válidas que las ponderaciones, las utilizaremos en nuestro ejemplo.

Cuando se emplea el enfoque de las cargas discriminantes, necesitamos conocer qué variables son discriminantes dignas de mención. En el análisis discriminante simultáneo, todas las variables entraron en la función, y generalmente todas las variables que mostraran una carga de ±0,30

TABLA 5.13. Resumen de las medidas interpretativas para el análisis discriminante de dos grupos

	Ponderaciones estandarizadas	*Cargas discriminantes*		*Ratio F univariante*	
Variables independientes	*Valor*	*Valor*	*Rango*	*Valor*	*Rango*
X_1 rapidez del envío	0,447	0,647	2	48,992	2
X_2 nivel de precios	NI	−0,433	4	8,453	5
X_3 flexibilidad de precios	0,472	0,522	3	31,881	3
X_4 imagen del fabricante	NI	−0,168	5	0,822	7
X_5 servicio global	NI	0,155	6	10,576	4
X_6 imagen del personal de ventas	NI	−0,145	7	1,620	6
X_7 calidad del producto	−0,659	−0,705	1	58,176	1

NI: No incluida en la solución por etapas.

o mayor se consideraban relevantes. Mediante los procedimientos por etapas, la determinación es más fácil de realizar porque los criterios especificados con esta técnica impiden que entren variables no significativas en la función. Sin embargo, la multicolinealidad y otros factores pueden impedir que una variable entre en la ecuación, pero esto no significa necesariamente que no cuente con una efecto sustancial. Esto puede determinarse evaluando las cargas discriminantes. Las cargas de las tres variables que entraron en la función discriminante son las tres mayores y las tres exceden el $\pm 0{,}30$, junto con X_2 (el nivel de precios). Para comprender los factores que distinguen entre estos dos grupos, el investigador debería considerar estas cuatro variables.

El investigador generalmente está interesado en interpretaciones de las variables individuales que cuenten con significación estadística y práctica. Tales interpretaciones se realizan identificando las variables que son significativas en la función discriminante, y comprendiendo lo que indican las diferentes medias de los grupos. Por ejemplo, para todas las variables de este modelo, puntuaciones mayores indican percepciones más favorables de HATCO sobre ese atributo (para más detalle, véase Capítulo 1). Mediante la Tabla 5.13 podemos emplear las cargas discriminantes (matriz de información de estructura) y los valores univariantes de la F para determinar cómo se clasifican estas variables en función de sus valores discriminantes. Ambas medidas presentan un alto grado de correspondencia. De las tres variables de la función, X_7 es la que más discrimina y X_3 la que menos. X_6 no se incluyó en el modelo, incluso aunque era equiparable a X_1 en capacidad discriminatoria en la tercera etapa del proceso de estimación, debido a su colinealidad con las variables ya incluidas, concretamente X_1, reducida la capacidad discriminatoria que podría ofrecer. Volviendo a la Tabla 5.5, nótese que dos de las tres variables (X_1, velocidad en el reparto y X_3, flexibilidad en el precio) tienen medias mayores en aquellas empresas que emplean el enfoque del valor total, significando esto que tienen percepciones más favorables de HATCO que las empresas que emplean la especificación de la compra. Solamente en la calidad del producto (X_7) la media de las empresas que utilizan la especificación de la compra es mayor. Se puede concluir que las empresas que emplean en enfoque de análisis del valor total utilizan un mayor conjunto de factores, mientras que las de especificación de la compra se centran en la calidad del producto.

Sexto paso: Validación de los resultados

La última etapa se centra en la validez externa e interna de la función discriminante. El primer método de validación es por medio del uso de la ampliación de la muestra y la valoración de su precisión predictiva. De esta manera, la validez se establece si la función discriminante actúa a un nivel aceptable de significación para clasificar las observaciones que no fueron empleadas en el proceso de estimación. Si la ampliación de la muestra está formada a partir de la muestra original, este enfoque presenta validación interna. Si la ampliación de la muestra proviene de otra muestra distinta, quizá de otra población o segmento de población, se permite la validación externa de los resultados discriminantes.

En nuestro ejemplo, la ampliación de la muestra proviene de la muestra original. Por ello, los niveles aceptables de las medidas de la precisión predictiva encontradas en la ampliación de la muestra sí establecen validación interna. Además, el análisis de los casos correctamente clasificados frente a los mal clasificados permite una mejor comprensión de las predicciones de grupo tanto en la muestra de análisis como en la ampliación de la muestra. A partir de los perfiles de los grupos tal como fueron mostrados en la Tabla 5.5, sabemos que, para las variables seleccionadas en el examen de las cargas discriminantes (X_1, X_2, X_3 y X_7), el grupo de especificación de compra tiene mejores percepciones de HATCO respecto de X_2 (nivel de precios) y de X_7 (calidad del producto), y el grupo de análisis del valor total tiene mayores percepciones sobre X_1 (rapidez en el envio) y de X_3 (flexibilidad en el precio). Así pues, se presentan los distintos perfiles para los dos grupos con los que llevar a cabo esfuerzos de marketing especializados. Se recomienda al investigador ampliar el proceso de validación por medio de un perfil más detallado de los grupos y el posible uso de muestras adicionales para establecer la validación externa.

Una visión gerencial de conjunto

El análisis discriminante de los clientes de HATCO basado en el tipo de estrategia de compra identificó un conjunto de diferencias perceptivas que ofrecen una distinción más clara y potente entre los dos grupos. Un resultado clave es que las diferencias se encontraron en un subconjunto de cuatro percepciones, permitiendo centrarse en variables clave, sin tener que considerar el conjunto entero. Las cuatro variables identificadas como discriminantes entre los grupos (mencionadas en orden de importancia) son X_7 (calidad del producto), X_1 (rapidez en el envío), X_3 (flexibilidad de precios) y X_2 (nivel de precios). Los resultados también indican que aquellas empresas que siguen un análisis de valor total tienen mayores percepciones de HATCO sobre las medidas de rapidez en el envío y flexibilidad de precios, quizá indicativas de una apreciación de las cualidades relativas de HATCO. El grupo de especificación de compra es más sensible a la calidad del producto y al nivel de precios, indicando quizá su interés por aspectos más fundamentales y objetivos de las transacciones. Por tanto, una vez que se identifica la estrategia de compra de una empresa y se conocen las variables clave, el gestor puede hacer uso de una estrategia que acentúa las percepciones positivas en su trato con los clientes para consolidar aún más su posición.

Los resultados, que son altamente significativos, otorgan al investigador la capacidad de identificar correctamente la estrategia de compra utilizada basada en estas percepciones más del 90 por ciento de las veces. Esto da confianza para el desarrollo de estrategias basadas en estos resultados, dado su mayor grado de consistencia. El análisis de las empresas mal clasificadas revela un pequeño número de empresas que parecen «fuera de lugar». Las empresas que utilizan el análisis del valor total recuerdan claramente a las empresas con especificación de compra en dos aspectos —la rapidez en el envío y la flexibilidad de precios. Igualmente, las empresas mal clasificadas del grupo de análisis del valor total difirieron en términos de sus percepciones. Estos resultados podrían indicar un verdadero error al determinar el tipo de estrategia de compra utilizada, pero más probablemente señalen a otra característica no considerada en el estudio y que podría dar alguna explicación a estas diferencias.

Un ejemplo ilustrativo de tres grupos

Para ilustrar la aplicación de un análisis discriminante de tres grupos, empleamos una vez más la base de datos de HATCO. En el ejemplo anterior, estábamos interesados solamente en la discriminación entre dos grupos, y fuimos capaces de desarrollar una función discriminante única y una única puntuación de corte para dividir los dos grupos. En este ejemplo, es necesario construir dos funciones discriminantes diferentes para diferenciar entre tres grupos. La primera función separa un grupo de los otros dos, y la segunda separa a los restantes dos grupos. Al igual que en el primer ejemplo, se estudiarán cada uno de los seis pasos en que consiste el proceso de construcción del modelo.

Primer paso: Objetivos del análisis discriminante

El objetivo de HATCO en este estudio es determinar la relación entre las percepciones que las empresas tienen de HATCO y el tipo de situación de compra a la que habitualmente se enfrentan. Las empresas que habitualmente tratan con HATCO en diferentes situaciones de compra pueden ver y evaluar a HATCO de forma diferente. El modelo discriminante resultante, como el modelo de dos grupos visto anteriormente, permite una determinación precisa de las percepciones mantenidas por las empresas para cada tipo de situación de compra. A partir de esta información, HATCO puede desarrollar estrategias con objetivos para cada situación de compra que acentúen los puntos fuertes detectados.

Segundo paso: Diseño de la investigación del análisis discriminante

Para contrastar esta relación, se utiliza un análisis discriminante empleando X_{14} como la variable dependiente y las percepciones de HATCO de estas empresas (X_1 a X_7) como las variables independientes. Nótese que X_{14} difiere de la variable dependiente con respecto al ejemplo de dos grupos en que cuenta con tres categorías en las que se clasifica una empresa según la situación de compra (tarea nueva, recompra modificada o recompra directa) que más a menudo se utiliza con HATCO. El tamaño muestral de 100, que de nuevo se dividirá en muestra de análisis y ampliada, da lugar a un adecuado caso de variable ratio independiente (de 9 a 1). En la muestra de análisis, sólo un grupo, con 15 observaciones, cae por debajo del nivel recomendado de 20 casos por grupo. Aunque el tamaño del grupo excediera de 20 si se utilizara la muestra entera en la fase de análisis, la necesidad de validar los resultados obliga a ampliar la muestra. El análisis se realiza prestando atención a la clasificación e interpretación de este grupo.

Tercer paso: Supuestos del análisis discriminante

Al igual que se mencionó en el ejemplo de dos grupos, los supuestos de normalidad, linealidad y colinealidad de las variables independientes ya se han tratado en profundidad en el Capítulo 2. El análisis llevado a cabo en el Capítulo 2 indicó que las variables independientes cumplían estos supuestos a niveles adecuados para permitir continuar con el análisis sin tratamientos adicionales. El supuesto restante, la igualdad de las matrices de dispersión o de varianzas covarianzas, también se analizó en el Capítulo 2. El contraste *M* de Box contrasta la igualdad de las matrices de dispersión de las variables independientes entre los tres grupos (categorías). El contraste señaló diferencias a un nivel de significación del 0,01. Aunque en muchos ejemplos este nivel de significación necesitará algún tipo de tratamiento, la sensibilidad del contraste M de Box al tamaño muestral y a otras características de las variables independientes lo hacen un contraste muy flexible. Por ello, se juzga que el contraste estadístico no ofrece pruebas de que las matrices de dispersión sean suficientemente diferentes como para requerir alguna corrección, y el análisis pueda continuar. Empleamos estimaciones de la varianza por grupos (*pooled*) en este análisis para permitir el uso del procedimiento de validación cruzada.

Cuarto paso: Estimación de la función discriminante y valoración del ajuste global

Al igual que en el ejemplo anterior, empezamos nuestro análisis repasando las medias y las desviaciones típicas de los grupos para ver si los grupos son significativamente diferentes en lo respectivo a una sola variable. La Tabla 5.14 refleja las medias y las desviaciones típicas de los grupos, la lambda de Wilks y los ratios *F* univariantes (ANOVAs simples) para cada variable independiente. El repaso de los niveles de significación de las variables individuales revela que sobre una base univariante, todas las variables, excepto X_4 y X_6 presentan diferencias significativas entre las medias de los grupos. Aunque la inspección visual de las medias de los grupos pueda revelar algo sobre las diferencias entre los grupos (por ejemplo, el 1 frente al 2 y al 3, el 1 y el 2 frente al 3, el 1 frente al 2 y frente al 3, etc.), desconocemos la significación estadística de cualquier comparación concreta, sólo que existen diferencias significativas globales. Esto resultará importante en el análisis discriminante de tres o más grupos en tanto va a construirse una función discriminante más con cada función, permitiendo discriminar entre conjuntos de grupos. En el ejemplo sencillo del comienzo del capítulo, una variable discriminaba entre el grupo 1 frente al 2 y al 3, mientras que otra discriminaba entre el grupo 2 frente al 3 y al 1. Ésta es una de las principales ventajas que aparecen en el uso del análisis discriminante.

TABLA 5.14. Estadísticos descriptivos de los grupos y contrastes de igualdad para el análisis discriminante de tres grupos

Variable dependiente[a]	*Medias de los grupos para las variables independientes*[b]							*Tamaño muestral*
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	
1: Nueva tarea	2,43	2,16	7,23	5,07	2,28	2,69	7,76	21
2: Recompra modificada	3,23	3,52	6,98	5,59	3,35	2,69	7,31	15
3: Recompra directa	4,64	1,93	9,18	5,41	3,30	2,80	5,92	24
Total	3,51	2,41	7,95	5,33	2,96	2,73	6,91	60
	Desviaciones estándar para las variables independientes[b]							
1: Nueva tarea	1,16	0,92	0,88	0,87	0,67	0,73	1,37	
2: Recompra modificada	1,13	1,38	1,34	1,12	0,64	0,77	1,75	
3: Recompra directa	1,02	0,89	0,70	1,39	0,37	0,94	1,41	
Total	1,46	1,21	1,38	1,16	0,75	0,82	1,68	
	Contrastes de igualdad de las medias de los grupos[c]							
Lambda de Wilks	0,550	0,709	0,461	0,967	0,546	0,996	0,754	
Ratio F univariante	23,346	11,674	33,362	0,961	23,692	0,126	9,293	
Nivel de significación	0,000	0,000	0,000	0,389	0,000	0,882	0,000	

[a] X_{14} = tipo de situación de compra
[b] X_1 = rapidez del envío; X_2 = nivel de precios; X_3 = flexibilidad de precios; X_4 = imagen del fabricante; X_5 = servicio global; X_6 = imagen del personal de ventas; X_7 = calidad del producto
[c] Lambda de Wilks (estadístico *U*) y ratio *F* univariante con 2 y 57 grados de libertad.

Estimación de la función discriminante

El procedimiento por etapas se lleva a cabo de la misma manera que en el ejemplo de dos grupos. Los datos de la Tabla 5.15 muestran que la primera variable que entra en el modelo es X_1 (rapidez en el envío). Como se hizo en el análisis de dos grupos, para que se considere su inclusión las variables han de tener en primer lugar un nivel de significación del 0,05 o por debajo. Luego, para las variables que cumplan este criterio, la variable con la mayor medida D^2 de Mahalanobis se selecciona para su inclusión en la función discriminante. El repaso de los niveles de significación revela que las variables no incluidas en el modelo tras la primera etapa (véase Tabla 5.15), todas excepto X_4, X_6 y X_7, tienen niveles de significación lo bastante bajos como para que se considere su inclusión en el modelo en las etapas posteriores.

La Tabla 5.16 detalla la segunda fase del procedimiento por etapas, que añade X_2 (el nivel de precios) a la función discriminante. La discriminación entre grupos ha crecido, como se refleja en el valor más bajo de la lambda de Wilks y el aumento de la D^2 (de 0,526 a 2,718). De las variables que no están en la ecuación, sólo X_3 (flexibilidad del precio) alcanza el nivel de significación necesario para ser considerada. Cuando se añade X_3 al modelo en la tercera etapa (véase la Tabla 5.17) no pueden considerarse para su inclusión otras variables porque todos sus niveles de significación quedan por encima del 0,05. Sin más variables que añadir finaliza el procedimiento de estimación.

La información contenida en la Tabla 5.18 resume las etapas del análisis discriminante de tres grupos. Sólo las variables X_1, X_2 y X_3 entraron en la función discriminante. La discriminación creció con la adición de cada variable, alcanzando en la tercera etapa una capacidad sustancial pa-ra discriminar entre los grupos. Comparando la lambda de Wilks final del análisis discriminante (0,222) con la lambda de Wilks (0,461) del mejor resultado con una única variable, X_3, observamos una notable mejora al utilizar la función discriminante en lugar de una única variable.

TABLA 5.15. Resultados del primer paso del modelo de análisis discriminante de tres grupos por pasos

PRIMER PASO: X_1 (RAPIDEZ EN EL ENVÍO) INCLUIDA EN EL ANÁLISIS

ESTADÍSTICOS RESUMEN

		Grados de libertad			*Significación*	*Entre los grupos*
Lambda de Wilks	0,550	1	2	57		
F equivalente	23,346		2	57	0,000	
Mínima D^2	0,526					1 y 2
F equivalente	4,606		1	57	0,036	

VARIABLES EN EL ANÁLISIS DESPUÉS DEL PRIMER PASO

		F para eliminar	
Variables	*Tolerancia*	*Valor*	*Significación*
X_1 rapidez del envío	1,00	23,346	0,000

VARIABLES FUERA DEL ANÁLISIS DESPUÉS DEL PRIMER PASO

			F para entrar			
Variables	*Tolerancia*	*Tolerancia mínima*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_2 nivel de precios	0,766	0,766	14,609	0,000	2,718	2 y 3
X_3 flexibilidad de precios	0,982	0,982	17,273	0,000	0,660	1 y 2
X_4 imagen del fabricante	0,998	0,998	0,690	0,506	0,698	1 y 2
X_5 servicio global	0,704	0,704	13,243	0,000	2,537	2 y 3
X_6 imagen del personal de ventas	0,991	0,991	0,020	0,980	0,531	1 y 2
X_7 calidad del producto	0,869	0,869	1,142	0,326	0,529	1 y 2

CONTRASTACIÓN DE LA SIGNIFICACIÓN DE LAS DIFERENCIAS DE GRUPO DESPUÉS DEL PRIMER PASO[a]

	Grupo 1: Nueva tarea	*Grupo 2: Recompra modificada*
Grupo 2: Recompra modificada	4,606 (0,036)	
Grupo 3: Recompra directa	45,335 (0,000)	15,274 (0,000)

[a] El estadístico *F* y el nivel de significación (en paréntesis) entre pares grupos después del primer paso. Cada estadístico *F* tiene 1 y 57 grados de libertad.

Significación estadística

La Tabla 5.18 también contiene la influencia global de las funciones discriminantes. Nótese que las funciones son estadísticamente significativas, medidas por el estadístico chi-cuadrado, y que la primera función explica el 78.4 por ciento de la varianza explicada por las dos funciones. Después de extraer la primera función, se recalcula la chi-cuadrado.

Los resultados muestran que existen diferencias significativas en la varianza restante. Si se utilizan más grupos en el modelo (por ejemplo, un análisis discriminante de cuatro grupos), son posibles funciones discriminantes canónicas adicionales, y se recalcularía el estadístico chi-cuadrado sobre la varianza residual para contrastar la presencia de diferencias significativas hasta que se extrajera el máximo número de funciones discriminantes canónicas (máximo número de funciones discriminantes = número de grupos – 1). La cantidad total de varianza explicada por la primera función es $0{,}812^2$, o el 65,9 por ciento. La siguiente función explica el $0{,}589^2$, o el 34,7 por ciento de la varianza restante (34,1 por ciento). Además, la varianza total explicada por ambas funciones es del 65,9 por ciento + (34,7 por ciento × 0,341), o el 77,7 por ciento de la variación total de la variable dependiente.

TABLA 5.16. Resultados del segundo paso del modelo de análisis discriminante de tres grupos por pasos

SEGUNDO PASO: X_2 (NIVEL DE PRECIOS) INCLUIDA EN EL ANÁLISIS

ESTADÍSTICOS RESUMEN

		Grados de libertad			*Significación*	*Entre los grupos*
Lambda de Wilks	0,361	2	2	57		
F equivalente	18,587		4	112	0,000	
Mínima D^2	2,718					2 y 3
F equivalente	12,325		2	56	0,000	

VARIABLES EN EL ANÁLISIS DESPUÉS DEL SEGUNDO PASO

		F para eliminar			
Variables	*Tolerancia*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_1 rapidez en el envío	0,766	26,988	0,000	0,046	1 y 3
X_2 nivel de precios	0,766	14,609	0,000	0,526	1 y 2

VARIABLES FUERA DEL ANÁLISIS DESPUÉS DEL SEGUNDO PASO

			F para entrar			
Variables	*Tolerancia*	*Tolerancia mínima*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_3 flexibilidad de precios	0,878	0,684	17,162	0,000	4,201	1 y 2
X_4 imagen del fabricante	0,963	0,739	0,009	0,991	2,719	2 y 3
X_5 servicio global	0,017	0,017	0,846	0,435	2,975	2 y 3
X_6 imagen del personal de ventas	0,952	0,736	0,534	0,589	2,766	2 y 3
X_7 calidad del producto	0,786	0,693	2,140	0,127	2,755	2 y 3

CONTRASTACIÓN DE LA SIGNIFICACIÓN DE LAS DIFERENCIAS DE GRUPO DESPUÉS DEL SEGUNDO PASO[a]

	Grupo 1: Nueva tarea	*Grupo 2: Recompra modificada*
Grupo 2: Recompra modificada	17,749 (0,000)	
Grupo 3: Recompra directa	26,405 (0,000)	12,325 (0,000)

[a] El estadístico *F* y el nivel de significación (en paréntesis) entre pares grupos después del segundo paso. Cada estadístico *F* tiene 2 y 56 grados de libertad.

Valoración del ajuste global

Las funciones discriminantes estimadas son combinaciones lineales similares a la recta de regresión (esto es, son combinaciones lineales de variables). Al igual que una recta de regresión trata de explicar la máxima cantidad de variación de su variable dependiente, estas combinaciones lineales intentan explicar las variaciones o diferencias en la variable categórica dependiente. La primera función discriminante se construye para explicar la mayor cantidad de variación (diferencia) en los grupos discriminantes. La segunda función discriminante, que es ortogonal e independiente de la primera, explica el mayor porcentaje de la varianza restante (residual) tras haber eliminado la varianza de la primera función.

Cálculo de las puntuaciones Z discriminantes La Tabla 5.18 también muestra los coeficientes de las funciones discriminantes (ponderaciones) y la matriz de estructura de las cargas discriminantes. Se presentan los valores rotados y no rotados. La rotación de las funciones discriminantes facilita la interpretación de la misma forma que se simplificaba la de los factores por

TABLA 5.17. Resultados del tercer paso del modelo de análisis discriminante de tres grupos por pasos

TERCER PASO: X_3 (FLEXIBILIDAD DE PRECIOS) INCLUIDA EN EL ANÁLISIS

ESTADÍSTICOS RESUMEN

		Grados de libertad			*Significación*	*Entre los grupos*
Lambda de Wilks	0,222	3	2	57		
F equivalente	20,540		6	110	0,000	
Mínima D^2	4,201					1 y 2
F equivalente	11,824		3	55	0,000	

VARIABLES EN EL ANÁLISIS DESPUÉS DEL TERCER PASO

		F para eliminar			
Variables	*Tolerancia*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_1 rapidez en el envío	0,764	.	0,000	1,757	1 y 2
X_2 nivel de precios	0,684	.	0,000	0,660	1 y 2
X_3 flexibilidad de precios	0,878	.	0,000	2,718	1 y 3

VARIABLES FUERA DEL ANÁLISIS DESPUÉS DEL TERCER PASO

			F para entrar			
Variables	*Tolerancia*	*Tolerancia mínima*	*Valor*	*Significación*	D^2	*Entre los grupos*
X_4 imagen del fabricante	0,962	0,665	.	0,982	4,206	1 y 2
X_5 servicio global	0,017	0,017	.	0,585	4,376	1 y 2
X_6 imagen del personal de ventas	0,944	0,652	.	0,535	4,432	1 y 2
X_7 calidad del producto	0,780	0,615	.	0,121	4,678	1 y 2

CONTRASTACIÓN DE LA SIGNIFICACIÓN DE LAS DIFERENCIAS DE GRUPO DESPUÉS DEL TERCER PASO[a]

	Grupo 1: Nueva tarea	*Grupo 2: Recompra modificada*
Grupo 2: Recompra modificada	11,824	
	(0,000)	
Grupo 3: Recompra directa	34,432	18,934
	(0,000)	(0,000)

[a] El estadístico *F* y el nivel de significación (en paréntesis) entre pares grupos después del tercer paso. Cada estadístico *F* tiene 3 y 55 grados de libertad.

la rotación (véase Capítulo 3 para una discusión más detallada de la rotación). Examinaremos los valores rotados y no rotados con más detalle en la quinta etapa. Las funciones de clasificación se utilizan para hacer predicciones de pertenencia al grupo, y la sección de medias de los grupos de la Tabla 5.18 perfila cada grupo mediante las puntuaciones *Z* discriminantes medias para cada función.

Debido a que éste es un modelo de análisis discriminante de tres grupos, se calculan dos funciones discriminantes para discriminar entre los tres grupos. Los valores para cada caso entran en el procedimiento discriminante y se formulan combinaciones lineales. Las funciones discriminantes se basan sólo en las variables incluidas en el modelo discriminante (X_1, X_2 y X_3). En los ejemplos de este capítulo, las puntuaciones *Z* discriminantes se basan en los coefientes rotados.

Después de calcular las combinaciones lineales, el procedimiento correlaciona las siete variables independientes con las funciones discriminantes para construir una matriz de estructura (cargas). Este procedimiento nos permite ver dónde tendría lugar la discriminación si las siete variables se incluyeran en el modelo (esto es, si ninguna fuera excluida por multicolinealidad o ausencia de significatividad estadística). Gran parte de esta exposición está basada en conceptos presentados en los capítulos sobre correlación canónica (Capítulo 8) y análisis de factores (Capítulo 3).

TABLA 5.18. Resultados multivariantes del análisis discriminante de tres grupos

TABLA RESUMEN

	Acción		*Lambda de Wilks*		*Mínima D^2*		
Etapas	*Entró*	*Se eliminó*	*Valor*	*Significación*	*Valor*	*Significación*	*Entre los grupos*
1	X_1 rapidez en el envío		0,550	0,000	0,526	0,036	1 y 2
2	X_2 nivel de precios		0,361	0,000	2,718	0,000	2 y 3
3	X_3 flexibilidad de precios		0,222	0,000	4,201	0,000	1 y 2

FUNCIONES DISCRIMINANTES CANÓNICAS

		Porcentaje de varianza							
Función	*Autovalor*	*Función*	*Acumulado*	*Correlación canónica*	*Tras la función*	*Lambda de Wilks*	*Chi-cuadrado*	*df*	*Significación*
					0	0,222	84,177	6	0,000
1	1,935	78,4	78,4	0,812	1	0,653	23,879	2	0,000
2	0,532	21,6	100,0	0,589					

COEFICIENTES DE LAS FUNCIONES DISCRIMINANTES CANÓNICAS

	Estandarizada				*No estandarizada*			
	No rotada		*Rotada*		*No rotada*		*Rotada*	
Variables independientes	*Función* 1	*Función* 2	*Función* 1	*Función* 2	*Función* 1	*Función* 2	*Función* 1	*Función* 2
X_1 rapidez del envío	0,785	0,559	0,834	0,088	0,713	0,508	0,420	0,768
X_2 nivel de precios	0,495	0,995	0,011	1,111	0,476	0,957	0,011	1,068
X_3 flexibilidad de precios	0,788	−0,285	0,462	0,845	0,825	−0,298	0,872	0,092
Constante					−10,207	−1,720	−8,434	−6,001

MATRIZ DE ESTRUCTURA NO ROTADA Y ROTADA DE LAS FUNCIONES DISCRIMINANTES[b]

	Cargas de las funciones discriminantes no rotadas		*Cargas de las funciones discriminantes rotadas*	
Variables independientes	*Función* 1	*Función* 2	*Función* 1	*Función* 2
X_1 rapidez del envío	0,650[c]	0,039	0,568[c]	0,319
X_2 nivel de precios	−0,159	0,824[c]	−0,502	0,672[c]
X_3 flexibilidad de precios	0,721[c]	−0,566	0,891[c]	−0,186
X_4 imagen del fabricante	0,046	0,188[c]	−0,041	0,190[c]
X_5 servicio global	0,509	0,817[c]	0,101	0,957[c]
X_6 imagen del personal de ventas	0,161	0,171[c]	0,071	0,224[c]
X_7 calidad del producto	−0,129	0,243[c]	−0,222[c]	0,163

MEDIAS DE LOS GRUPOS (CENTROIDES) NO ROTADAS Y ROTADAS DE LAS FUNCIONES DISCRIMINANTES CANÓNICAS

	Centroides de grupo no rotadas		*Centroides de grupo rotadas*	
Grupo	*Función* 1	*Función* 2	*Función* 1	*Función* 2
Nueva tarea	−1,482	−0,579	−1,081	−1,167
Recompra modificada	−0,473	1,206	−0,952	0,879
Recompra directa	1,592	−0,247	1,541	0,472

(Continúa)

TABLA 5.18. (Continuación)

COEFICIENTES DE LA FUNCIÓN DE CLASIFICACIÓN

	X_{14} *Tipo de situación de compra*		
Variables independientes	*Grupo 1: Nueva tarea*	*Grupo 2: Recompra modificada*	*Grupo 3: Recompra directa*
X_1 rapidez en el envío	4,207	5,833	6,568
X_2 nivel de precios	7,356	9,543	9,135
X_3 flexibilidad de precios	10,060	10,360	12,498
Constante	−50,478	−63,750	−82,323

[a] Señala la función discriminante canónica.
[b] Correlaciones entre grupos entre variables discriminantes y funciones discriminantes canónicas.
[c] Denota la correlación absoluta mayor entre cada variable y cualquier función discriminante.

Evaluación de las diferencias de grupo Incluso aunque ambas funciones discriminantes fueran estadísticamente significativas, el investigador debe siempre asegurarse de que la función discriminante suministre diferencias entre todos los grupos. Es posible tener funciones estadísticamente significativas, pero tener al menos un par de grupos que no sean estadísticamente diferentes (esto es, sin discriminación entre ambos). Este problema llega a prevalecer conforme crece el número de grupos y/o el número de grupos pequeños incluido en el análisis. La última sección de la Tabla 5.17 presenta los contrastes de significación para diferencias de grupos entre cada par de grupos (por ejemplo, el grupo 1 frente al grupo 2, el grupo 1 frente al grupo 3, etc.). Todos los pares de grupos mostraron diferencias estadísticamente significativas, denotando que las funciones discriminantes separaban no sólo en un sentido global, sino también para cada grupo. También examinamos gráficamente los centroides de grupo en la última sección.

Valoración de la precisión predictiva de pertenencia al grupo La última etapa en la valoración del ajuste global del modelo consiste en determinar el nivel de precisión predictiva de la(s) función(es) discriminante(s). Esta determinación se realiza de la misma forma que en el modelo discriminante de dos grupos por medio del examen de las matrices de clasificación. La Tabla 5.19 muestra que las dos funciones discriminantes en combinación logran un alto grado de precisión clasificatoria. El ratio de aciertos para la muestra de análisis es del 81,7 por ciento, mientras que para la ampliación de la muestra es del 75,0 por ciento. Estos resultados señalan un sesgo al alza que existe probablemente cuando no se emplea la ampliación de la muestra. Aunque ambos ratios de aciertos son altos, deben compararse con los criterios de aleatoriedad proporcional y de aleatoriedad máxima para valorar su efectividad «real». Una tercera matriz de clasificación para el procedimiento de validación cruzada se discute en el sexto paso.

El criterio de aleatoriedad máxima es simplemente el ratio que se obtiene si asignamos a todas las observaciones del grupo la más alta probabilidad de ocurrencia. En la presente muestra de 100 observaciones, 34 se tomaron para el grupo 1, 32 para el grupo 2 y 34 para el grupo 3. Con esta información, podemos comprobar que la probabilidad más alta sería del 34 por ciento (grupos 1 o 3). Por tanto, si nos basamos en el criterio de máxima aleatoriedad nuestro modelo es muy bueno. El criterio de aleatoriedad proporcional se calcula elevando al cuadrado las proporciones de cada grupo, con un valor calculado del 33,36 por ciento. Dado que C_{MAX} es mayor que C_{PRO}, emplearemos el criterio de máxima aleatoriedad. Si establecemos el umbral de un 25 por ciento mayor que el valor del criterio, el ratio de aciertos debe exceder el 42,5 por ciento (34 por ciento × 1,25). Los ratios de aciertos del 83,3 por ciento (muestra de análisis) y del 75,0 por ciento (ampliación de la muestra) exceden ambos este criterio sustancialmente, pero deberíamos también evaluar la tasa de discriminación para grupos individuales.

Tanto para la muestra de análisis como para la ampliación de la muestra, los grupos 1 y 3 tienen ratios de aciertos que exceden con mucho el umbral del 42,5 por ciento. El grupo 2, sin embargo, tiene un ratio de aciertos del 69 por ciento y del 47,5 por ciento en las muestras de análisis

TABLA 5.19. Matrices de clasificación para el análisis discriminante de tres grupos

RESULTADOS DE LA CLASIFICACIÓN: MUESTRA DE ANÁLISIS[a]

		Pertenencia al grupo predicha		
Grupo real	*Número de casos*	*Nueva tarea*	*Recompra modificada*	*Recompra directa*
Nueva tarea	21	16 76,2%	3 14,3%	2 9,5%
Recompra modificada	15	3 20,0%	9 60,0%	3 20,0%
Recompra directa	24	0 0%	0 0%	24 100%
Número de casos		19	12	29

RESULTADOS DE LA CLASIFICACIÓN: AMPLIACIÓN DE LA MUESTRA[b]

		Pertenencia al grupo predicha		
Grupo real	*Número de casos*	*Nueva tarea*	*Recompra modificada*	*Recompra directa*
Nueva tarea	13	12 92,3%	1 7,7%	0 0%
Recompra modificada	17	4 23,5%	8 47,1%	5 29,4%
Recompra directa	10	0 0%	0 0%	10 100,0%
Número de casos		16	9	15

RESULTADOS DE LA CLASIFICACIÓN: VALIDACIÓN CRUZADA DE LA MUESTRA DE ANÁLISIS[c]

		Pertenencia al grupo predicha		
Grupo real	*Número de casos*	*Nueva tarea*	*Recompra modificada*	*Recompra directa*
Nueva tarea	21	16 76,2%	3 14,3%	2 9,5%
Recompra modificada	15	3 20,0%	9 60,0%	3 20,0%
Recompra directa	24	1 4,2%	0 0%	23 95,8%
Número de casos		20	12	28

[a] Porcentaje de los casos de la muestra de análisis correctamente clasificados: 81,7 por ciento [(16 + 9 + 24)/60 = 0,817].
[b] Porcentaje de los casos de la ampliación de la muestra correctamente clasificados: 75,0 por ciento [(12 + 8 + 10)/40 = = 0,750].
[c] Porcentaje de los casos de validación cruzada correctamente clasificados: 80,0 por ciento [(16 + 9 + 23)/60 = 0,800].

y ampliada, respectivamente. Por tanto, aunque los tres grupos sí exceden el valor del umbral, las predicciones del grupo 2 deberían mejorarse en lo posible, probablemente añadiendo variables independientes o mediante un repaso de la clasificación de las empresas en este grupo para identificar las características de este grupo no representadas en la función discriminante. Cuando se haya terminado, el investigador puede concluir confiadamente que el modelo discriminante es válido y tiene niveles adecuados de significatividad estadística y práctica para todos los grupos.

La última medida de precisión clasificatoria es la Q de Press, calculada tanto para la muestra de análisis como para la ampliación de la muestra. Contrasta la significación estadística de que la precisión clasificatoria es mejor que la clasificación meramente aleatoria. El valor calculado para la muestra de análisis es

$$Q \text{ de Press}_{\text{muestra de análisis}} = \frac{[60 - (49 \times 3)]^2}{60(3 - 1)} = 63{,}1$$

El valor calculado para la ampliación de la muestra es

$$Q \text{ de Press}_{\text{ampliación de la muestra}} = \frac{[40 - (30 \times 3)]^2}{40(3 - 1)} = 31{,}2$$

Como el valor crítico a un nivel de significación de 0,01 es de 6,63, el análisis discriminante se puede considerar como un mejor predictor de la pertenencia a un grupo que la aleatoriedad, teniendo en cuenta tanto ésta como las otras medidas de precisión clasificatoria.

Diagnóstico por casos

La Tabla 5.20 contiene datos de clasificación adicionales para cada caso individual. El número de observaciones aparece reflejado en el lado izquierdo de la tabla, que también contiene notación sobre los casos incluidos en la ampliación de la muestra. En la columna «grupo real», un 1 indica el grupo 1 (tarea nueva), un 2 indica el grupo 2 (recompra modificada) y un 3 indica el grupo 3 (recompra directa). Un asterisco al lado de los números señala que una observación determinada fue mal clasificada por la función discriminante. La columna «grupo de probabilidad mayor» muestra la asignación más probable de una observación empleando la función de clasificación discriminante, y la columna «segunda mayor probabilidad» muestra la segunda mejor probabilidad utilizando las funciones de clasificación discriminantes. También se suministran las puntuaciones discriminantes para cada observación sobre cada función. Cuando existe una función discriminante (dos grupos), la clasificación de los casos se basa en los valores de la única función; las funciones múltiples se emplean para tres o más grupos. Los casos mal clasificados se indican cuando difieren las predicciones con la verdadera pertenencia al grupo. Los valores de la probabilidad permiten al investigador valorar la importancia de la clasificación errónea. Como ejemplo, supongamos dos situaciones: en la primera el grupo incorrecto tiene una probabilidad de 0,53 y el correcto de 0,46; en una segunda situación el grupo incorrecto tiene una probabilidad de 0,76 y el correcto de 0,22. En ambas situaciones hubo una clasificación errónea, pero la medida o magnitud varía ampliamente. El investigador debería evaluar la medida de la clasificación errónea en cada caso.

Clasificación errónea de casos individuales Usando la información de la Tabla 5.20, los casos mal clasificados en cada grupo pueden identificarse y compararse con los casos correctamente clasificados de ese grupo para identificar diferencias. La Tabla 5.21 contiene un análisis de las diferencias para el análisis discriminante de tres grupos, que revela patrones distintos. Los casos mal clasificados parecen bastante uniformes en sus diferencias sugiriendo la posibilidad de un cuarto grupo. Primero, el grupo de nueva tarea cuenta con percepciones mayores casi uniformemente que los correctamente clasificados, con diferencias significativas para cuatro variables (X_1, X_2, X_5 y X_6). Casi lo contrario sucede para el grupo de compra modificada —los casos mal clasificados son significativamente menores en sus percepciones de las seis variables (de X_2 a X_7, excepto para X_3). Combinados, estos casos mal clasificados parecen formar un grupo medio con percepciones entre los dos grupos existentes. Los investigadores de HATCO deberían usar este hallazgo para investigar si una variable adicional ayudaría a distinguir estos casos, o si la variable independiente captura inadecuadamente estos casos en la clasificación actual en tres grupos.

TABLA 5.20. Datos de la clasificación para el análisis discriminante de tres grupos

		Predicción de pertenencia al grupo									
		Mayor probabilidad:					*Segunda mayor probabilidad:*			*Puntuaciones discriminantes*	
Número de caso	*Grupo real*	*Grupo*	*P(D/G)*	*gl*	*P(G/D)*	D^2	*Grupo*	*P(G/D)*	D^2	*Función 1*	*Función 2*
1	1	1	0,851	2	0,949	0,324	2	0,038	6,10	−0,686	−1,577
2	1	1	0,534	2	0,869	1,254	2	0,130	4,373	−2,150	−0,835
3[u]	2	2	0,115	2	0,993	4,330	1	0,005	15,673	−1,978	2,689
4[u]	1	1	0,582	2	0,988	1,081	2	0,010	9,54	−1,095	−2,207
5	3	3	0,649	2	0,997	0,866	2	0,003	11,899	2,471	0,451
6	2	1*	0,636	2	0,603	0,905	2	0,332	1,428	−0,709	−0,291
7	1	3*	0,852	2	0,983	0,320	2	0,016	7,645	1,812	0,968
8	2	1*	0,190	2	0,528	3,322	2	0,471	2,879	−2,435	0,053
9[u]	3	3	0,814	2	0,992	0,411	2	0,007	9,284	2,094	0,796
10[u]	2	2	0,866	2	0,908	0,288	1	0,053	6,629	−1,046	1,407
11	1	1	0,360	2	0,840	2,043	3	0,127	6,083	0,268	−1,641
12	2	3*	0,894	2	0,918	0,225	2	0,053	4,975	1,166	0,180
13	1	1	0,602	2	0,918	1,014	3	0,046	7,289	−0,177	−1,611
14	1	3*	0,309	2	0,574	2,349	1	0,342	3,117	0,639	−0,767
15	3	3	0,690	2	0,993	0,742	2	0,004	10,821	2,191	−0,093
16[u]	3	3	0,705	2	0,947	0,701	1	0,032	7,193	1,478	−0,362
17	2	2	0,475	2	0,928	1,490	1	0,069	7,373	−2,074	1,360
18[u]	2	3*	0,497	2	0,604	1,398	2	0,285	1,959	0,361	0,393
19[u]	3	3	0,767	2	0,995	0,532	2	0,004	10,555	2,269	0,456
20	3	3	0,690	2	0,993	0,742	2	0,004	10,821	2,191	−0,093
21[u]	2	1*	0,261	2	0,794	2,687	3	0,178	5,943	0,464	−1,715
22[u]	1	1	0,246	2	0,936	2,809	3	0,053	8,833	0,239	−2,200
23	3	3	0,492	2	0,803	1,420	2	0,188	3,380	0,807	1,412
24	1	1	0,660	2	0,983	0,832	2	0,017	8,320	−1,565	−1,940
25	2	3*	0,942	2	0,942	0,120	2	0,038	5,575	1,313	0,211
26	3	3	0,529	2	0,638	1,273	2	0,277	2,004	0,413	0,501
27[u]	1	1	0,626	2	0,983	0,938	2	0,016	8,486	−1,652	−1,949
28	3	3	0,775	2	0,994	0,510	2	0,004	10,468	2,226	0,273
29	3	3	0,980	2	0,978	0,040	2	0,018	7,132	1,703	0,587
30[u]	2	2	0,634	2	0,952	0,913	1	0,037	8,093	−1,525	1,643
31	1	2*	0,528	2	0,596	1,276	1	0,399	2,753	−1,905	0,273
32	3	3	0,524	2	0,702	1,294	2	0,270	2,268	0,547	1,026
33	3	3	0,827	2	0,985	0,380	2	0,013	8,058	1,885	0,984
34[u]	1	2*	0,878	2	0,890	0,260	1	0,092	5,464	−1,383	1,151
35[u]	1	1	0,444	2	0,989	1,623	2	0,008	10,698	−0,698	−2,382
36	1	1	0,748	2	0,691	0,581	2	0,287	1,669	−1,10	−0,405
37[u]	2	2	0,541	2	0,950	1,230	1	0,044	8,037	−1,819	1,570
38[u]	3	3	0,260	2	0,779	2,691	1	0,197	5,178	1,195	−1,131
39	1	1	0,062	2	0,999	5,552	2	0,001	18,10	−2,392	−3,125
40[u]	1	1	0,734	2	0,955	0,619	2	0,044	6,101	−1,821	−1,433
41[u]	1	1	0,630	2	0,593	0,925	2	0,371	1,190	−0,970	−0,212
42	3	3	0,671	2	0,996	0,799	2	0,003	11,684	2,429	0,374
43	3	3	0,814	2	0,980	0,411	2	0,019	7,404	1,762	1,074
44[u]	2	3*	0,846	2	0,916	0,334	2	0,048	5,297	1,183	0,018
45	1	1	0,715	2	0,915	0,670	2	0,084	4,769	−1,896	−1,091
46[u]	3	3	0,626	2	0,984	0,937	2	0,015	8,341	1,893	1,373
47	3	3	0,518	2	0,949	1,315	1	0,038	7,497	1,612	−0,672
48	2	2	0,487	2	0,919	1,440	1	0,078	7,039	−2,079	1,291
49	3	3	0,603	2	0,963	1,012	1	0,025	8,044	1,682	−0,524
50	3	3	0,630	2	0,826	0,923	2	0,161	3,257	0,836	1,125
51	2	1*	0,351	2	0,840	2,096	3	0,128	6,124	0,281	−1,658

(Continúa)

TABLA 5.20. (Continuación)

		Predicción de pertenencia al grupo									
		Mayor probabilidad:					*Segunda mayor probabilidad:*			*Puntuaciones discriminantes*	
Número de caso	*Grupo real*	*Grupo*	*P(D/G)*	*gl*	*P(G/D)*	D^2	*Grupo*	*P(G/D)*	D^2	*Función 1*	*Función 2*
52	2	2	0,436	2	0,738	1,659	3	0,247	4,786	−0,137	1,876
53	2	2	0,307	2	0,963	2,360	3	0,030	10,246	−1,004	2,414
54	1	1	0,844	2	0,806	0,340	2	0,187	2,592	−1,387	−0,671
55[u]	1	1	0,177	2	0,542	3,458	3	0,423	4,221	0,761	−1,429
56[u]	2	1*	0,565	2	0,565	1,143	2	0,322	1,595	−0,471	−0,289
57[u]	2	2	0,043	2	0,720	6,290	3	0,279	9,130	0,127	3,143
58	3	3	0,370	2	0,992	1,989	2	0,007	10,836	2,236	1,699
59	3	3	0,271	2	0,347	2,609	1	0,337	2,403	0,022	−0,077
60[u]	2	2	0,801	2	0,791	0,445	3	0,157	4,623	−0,441	1,307
61	3	3	0,514	2	0,937	1,331	1	0,048	7,026	1,525	−0,681
62[u]	3	3	0,929	2	0,981	0,147	2	0,017	7,333	1,755	0,791
63[u]	3	3	0,429	2	0,904	1,693	1	0,077	6,338	1,413	−0,823
64[u]	2	2	0,384	2	0,825	1,914	1	0,173	5,715	−2,335	0,868
65	1	1	0,413	2	0,993	1,767	2	0,007	10,995	−1,670	−2,359
66[u]	2	3*	0,364	2	0,753	2,021	1	0,202	4,383	0,987	−0,837
67	3	3	0,953	2	0,953	0,097	2	0,041	5,470	1,383	0,740
68	2	2	0,483	2	0,746	1,455	1	0,251	4,305	−2,131	0,622
69[u]	3	3	0,934	2	0,932	0,137	2	0,057	4,787	1,226	0,667
70	2	2	0,690	2	0,534	0,743	3	0,278	2,987	−0,188	0,480
71	2	2	0,067	2	0,994	5,408	3	0,004	17,385	−1,717	3,074
72	3	3	0,587	2	0,893	1,066	2	0,102	4,467	1,095	1,404
73	3	3	0,887	2	0,978	0,241	2	0,013	7,969	1,738	0,023
74[u]	3	3	0,955	2	0,978	0,091	2	0,015	7,485	1,703	0,218
75[u]	1	1	0,930	2	0,880	0,144	2	0,114	3,555	−1,395	−0,954
76[u]	2	3*	0,984	2	0,969	0,033	2	0,026	6,342	1,556	0,652
77[u]	2	1*	0,314	2	0,740	2,316	3	0,218	5,028	0,412	−1,465
78[u]	3	3	0,959	2	0,958	0,083	2	0,028	6,189	1,442	0,202
79	1	1	0,296	2	0,995	2,434	2	0,005	12,485	−1,800	−2,552
80	3	3	0,784	2	0,886	0,486	2	0,061	4,897	1,063	−0,036
81	3	3	0,893	2	0,962	0,227	2	0,034	5,945	1,485	0,946
82	2	2	0,249	2	0,979	2,781	1	0,019	11,381	−2,159	2,030
83[u]	1	1	0,412	2	0,938	1,772	2	0,062	6,536	−2,411	−1,221
84	1	2*	0,888	2	0,645	0,238	1	0,279	2,585	−0,798	0,416
85[u]	2	3*	0,316	2	0,406	2,301	2	0,307	1,923	0,105	−0,019
86	1	1	0,574	2	0,570	1,110	2	0,410	1,096	−1,244	−0,127
87[u]	2	1*	0,407	2	0,459	1,799	2	0,276	2,145	−0,067	−0,289
88	1	2*	0,694	2	0,486	0,731	1	0,479	1,434	−1,059	0,030
89	1	1	0,783	2	0,971	0,488	2	0,023	7,302	−0,835	−1,821
90	3	3	0,939	2	0,965	0,126	2	0,031	6,074	1,512	0,826
91[u]	2	2	0,569	2	0,423	1,129	1	0,395	1,936	−0,339	0,011
92	2	3*	0,409	2	0,510	1,790	2	0,308	1,858	0,231	0,200
93	2	2	0,217	2	0,922	3,051	1	0,077	8,687	−2,638	1,335
94[u]	1	1	0,096	2	0,942	4,687	2	0,058	9,573	−3,245	−1,198
95	1	1	0,816	2	0,970	0,406	2	0,025	7,064	−0,903	−1,779
96	1	1	0,038	2	0,999	6,565	2	0,001	19,648	−2,582	−3,244
97	3	3	0,760	2	0,993	0,550	2	0,005	10,346	2,160	0,064
98[u]	1	1	0,148	2	0,912	3,820	2	0,088	7,824	−3,028	−0,996
99	1	1	0,862	2	0,795	0,296	2	0,194	2,447	−1,263	−0,654
100[u]	1	1	0,289	2	0,657	2,484	3	0,293	4,364	0,487	−1,332

*Clasificación incorrecta.
[u]Caso no seleccionado utilizado en las muestras ampliada o de validación.

TABLA 5.21. Perfil de las observaciones correcta e incorrectamente clasificadas en el análisis discriminante de tres grupos

Variable dependiente Grupo/variables de perfil	*Puntuaciones medias*			*Contraste t*	
	Correctamente clasificado	*Mal clasificado*	*Diferencia*	*Valor de t*	*Significación*
Nueva tarea (X_{14} = 1)	(N = 28)	(N = 6)			
X_1 rapidez del envío[a]	2,332	3,183	−0,851	−1,989	0,055
X_2 nivel de precios[a]	1,900	3,000	−1,100	−2,831	0,008
X_3 flexibilidad de precios[a]	7,039	7,583	−0,54	−1,189	0,243
X_4 imagen del fabricante	4,839	5,517	−0,677	−1,498	0,144
X_5 servicio global	2,043	3,100	−1,057	−4,897	0,000
X_6 imagen del personal de ventas	2,482	3,233	−0,751	−2,534	0,016
X_7 calidad del producto	7,629	7,550	0,079	0,216[b]	0,831
Recompra modificada (X_{14} = 2)	(N = 17)	(N = 15)			
X_1 rapidez del envío[a]	3,294	3,567	−0,273	−0,770	0,448
X_2 nivel de precios[a]	4,188	2,040	2,148	7,045[b]	0,000
X_3 flexibilidad de precios[a]	6,447	8,260	−1,813	−5,431	0,000
X_4 imagen del fabricante	5,953	5,127	0,826	2,463	0,020
X_5 servicio global	3,712	2,800	0,912	5,472	0,000
X_6 imagen del personal de ventas	3,012	2,373	0,638	2,875	0,007
X_7 calidad del producto	8,041	6,493	1,548	3,005	0,005
Recompra directa (X_{14} = 3)	(N = 34)	(N = 0)			
X_1 rapidez del envío[a]	4,365	ND	NA	NA	NA
X_2 nivel de precios[a]	1,865	ND	NA	NA	NA
X_3 flexibilidad de precios[a]	9,215	ND	NA	NA	NA
X_4 imagen del fabricante	5,238	ND	NA	NA	NA
X_5 servicio global	3,256	ND	NA	NA	NA
X_6 imagen del personal de ventas	2,671	ND	NA	NA	NA
X_7 calidad del producto	6,003	ND	NA	NA	NA

Nota: Contiene observaciones tanto de la muestra de análisis como de la ampliada.
ND: Sin datos.
NA: No aplicable por ausencia de observaciones.
[a] Variables incluidas en la función discriminante.
[b] Contraste t calculado con estimaciones de la varianza separadas, en lugar de con estimaciones de la varianza agrupadas (*pooled*), debido a que el contraste de Levene detectó diferencias significativas de las varianzas entre los grupos.

El análisis de los casos mal clasificados puede sustituirse por el examen gráfico de las observaciones individuales. La Figura 5.12 representa cada observación basada en sus dos puntuaciones Z discriminantes con un solapamiento de los mapas territoriales que representan las fronteras de las puntuaciones de corte para cada función. Viendo la dispersión de cada grupo alrededor de los centroides de grupo, vemos que el grupo de compra directa (grupo 3) está más concentrado, con un pequeño solapamiento con los otros dos grupos. La compacidad del grupo 3 se ve también en el hecho de que no hay malas clasifcaciones en este grupo. Esto contrasta con el grupo 1 (6 clasificaciones erróneas) y el grupo 2 (15 clasificaciones erróneas). La dispersión de los grupos 2 y 3 se representan en el gráfico y la posible existencia de un cuarto grupo puede verse en el centro derecha del gráfico. La representación gráfica es útil no sólo para identificar estos casos de clasificación errónea que pueden formar un nuevo grupo, sino también para identificar casos atípicos.

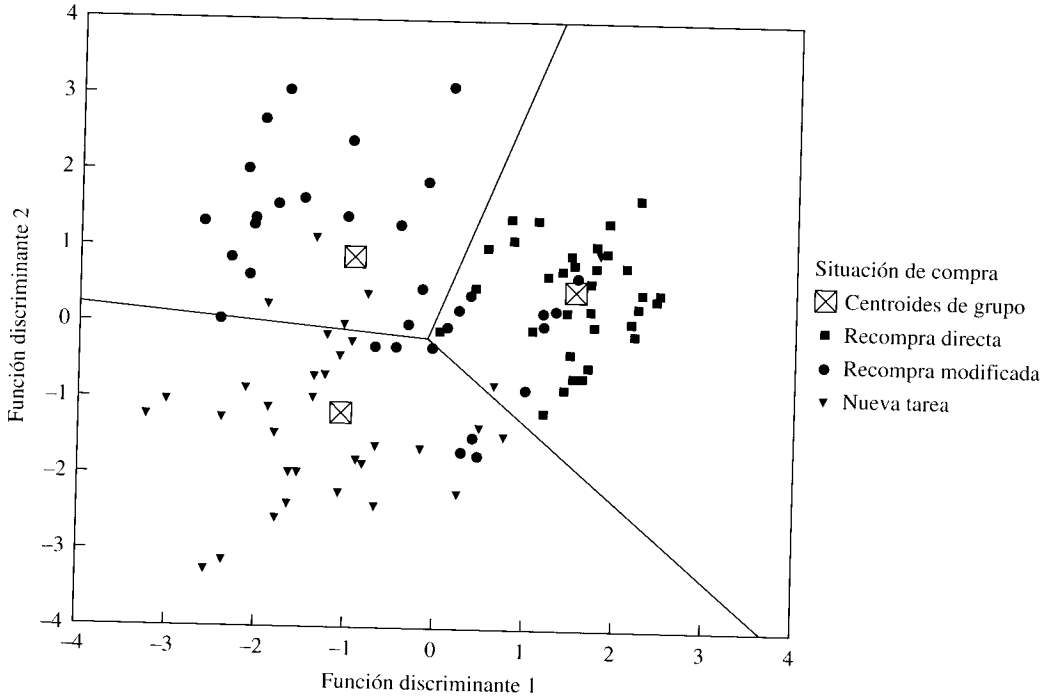


FIGURA 5.12. Mapa territorial y puntuaciones *Z* discriminantes rotadas.

Quinto paso: Interpretación de los resultados del análisis discriminante de tres grupos

La siguiente etapa del análisis discriminante es la interpretación de las funciones discriminantes. El primer paso consiste en revisar la rotación de las funciones con la finalidad de simplificar la interpretación. En el siguiente paso se examinan las contribuciones de las variables predictoras a cada función de forma individual (esto es, las cargas discriminantes). En el último paso se analizan los efectos acumulados de las variables independientes a lo largo de las funciones discriminantes múltiples con el índice de potencia. La representación gráfica puede también ayudar al investigador a comprender la posición relativa de cada grupo en esta solución bidimensional, y a interpretar las variables relevantes que determinan esta posición.

Rotación

Después de construir las funciones discriminantes, éstas pueden ser «rotadas» para redistribuir la varianza (este concepto está explicado de forma más completa en el Capítulo 3). Básicamente, la rotación mantiene la estructura original y la fiabilidad de los modelos discriminantes, haciendo al mismo tiempo más fácil una interpretación relevante. En este ejemplo, elegimos el procedimiento más ampliamente utilizado de la rotación VARIMAX. La rotación afecta a los coeficientes de la función y a las cargas discriminantes, además de al cálculo de las puntuaciones *Z* discriminantes y a los centroides de los grupos (véase Tabla 5.18). Examinando los coeficientes rotados frente a los no rotados o cargas, se revelan un conjunto de resultados algo más simplificados (esto es, cargas que tienden a separarse en valores altos frente a bajos en lugar de en el rango medio), pero en esta ocasión la rotación no es de ayuda como en otras ocasiones. Incluso, dado un análisis discriminante con dos o más funciones estimadas, la rotación puede ser una herramienta poderosa que debería considerarse siempre para mejorar la interpretabilidad de los resultados.

Valoración de la contribución de las variables predictoras

La primera tarea consiste en dibujar los centroides de los grupos para ofrecer una perspectiva global de las diferencias en los tres grupos. La Figura 5.12 representa los centroides de las primeras dos funciones junto con cada observación. Aunque existe algún solapamiento y alguna clasificación errónea, los tres grupos aparecen bien separados en la mayoría de los casos. Este gráfico muestra diferencias en los grupos debido a las funciones discriminantes, pero no ofrece una base para explicar estas diferencias en términos de las variables independientes.

Para valorar las contribuciones de los siete predictores, el investigador puede emplear un conjunto de medidas —cargas discriminantes, ratios F univariantes y el índice de potencia. Las técnicas incluidas en el uso de las cargas discriminantes y de los ratios F univariantes se comentaron en el ejemplo de dos grupos. Una técnica de interpretación adicional en una solución multifunción es el índice de potencia. La Tabla 5.22 ilustra el cálculo del índice de potencia para cada una de las variables predictoras. Las cargas representan la correlación entre la variable independiente y la puntuación Z discriminante. Así, la carga al cuadrado es la varianza de la variable independiente asociada con la función discriminante. Ponderando la varianza explicada de cada función por la capacidad discriminatoria relativa de las funciones y sumando entre funciones, el índice de potencia representa los efectos discriminantes totales a lo largo de todas las funciones discriminantes. Repasando los índices de potencia para las siete variables independientes se revela el hecho interesante de que X_5 cuenta con el segundo mayor valor, pero no fue incluido en la solución por etapas. ¿Cómo sucedió esto? Primero, recuérdese que la multicolinealidad puede afectar a la solución por etapas debido a la redundancia entre variables altamente colineales. X_5 era la segunda variable más discriminatoria en un sentido univariante (véase Tabla 5.14), pero la incorporación de X_3, que está correlacionada con X_5, dentro de la función discriminante hace a la capacidad discriminatoria de X_5 redundante (**tolerancia** de 0,017). Pero, ¿cómo es que X_5 tiene un alto valor de potencia incluso aunque no estaba en la solución discriminante final? Recuérdese que las cargas discriminantes se calcularon para todas las variables independientes, incluso si no estaban en la función discriminante. La finalidad está en ayudar en la interpretación en aquellos casos donde la multicolinealidad está presente.

La Tabla 5.23 presenta las tres medidas de interpretación preferidas para cada variable. Los resultados generalmente apoyan el análisis por etapas, aunque X_5 tiene un ratio F univariante y un índice de potencia importantes, no fue incluida debido a la colinealidad. Las otras variables no incluidas (X_2, X_4 y X_6) tienen todas cargas muy pequeñas, valores de la F no significativos y/o valores del índice de potencia bajos. Resulta de particular interés la interpretación de las dos dimensiones

TABLA 5.22. Cálculo de los índices de potencia para el análisis discriminante de tres grupos

	Función discriminante 1				*Función discriminante* 2				
Variables	*Carga (C)*	*Carga al cuadrado*	*Autovalor relativo*[a]	*Valor de potencia*[b]	*Carga (C)*	*Carga al cuadrado*	*Autovalor relativo*[a]	*Valor de potencia*[b]	*Índice de potencia*[c]
X_1 rapidez del envío	0,650	0,423	0,784	0,331	0,039	0,002	0,216	0,000	0,332
X_2 nivel de precios	−0,159	0,025	0,784	0,020	0,824	0,679	0,216	0,146	0,166
X_3 flexibilidad de precios	0,721	0,520	0,784	0,408	−0,566	0,320	0,216	0,069	0,477
X_4 imagen del fabricante	0,046	0,002	0,784	0,002	0,188	0,035	0,216	0,008	0,009
X_5 servicio global	0,509	0,259	0,784	0,203	0,817	0,667	0,216	0,144	0,347
X_6 imagen del personal de ventas	0,161	0,026	0,784	0,020	0,171	0,029	0,216	0,006	0,027
X_7 calidad del producto	−0,129	0,017	0,784	0,013	0,243	0,059	0,216	0,013	0,026

Nota: Se utilizan las cargas no rotadas para que se correspondan con los autovalores originales antes de la rotación.

[a] El autovalor relativo es el autovalor de la función discriminante dividido por la suma de los autovalores para todas las funciones significativas. En nuestro ejemplo, el autovalor relativo para la función 1 es [1,935/(1,935 + 0,532)] = 0,784.

[b] Valor de potencia = carga al cuadrado × autovalor relativo.

[c] Índice de potencia = valor de potencia de la función discriminante 1 + valor de potencia de la función discriminante 2.

TABLA 5.23. Resumen de las medidas interpretativas para el análisis discriminante de tres grupos

	Cargas de las funciones discriminantes rotadas			
Variables independientes	*Función 1*	*Función 2*	*Ratio univariante*	*Índice de potencia*
X_1 rapidez del envío	0,568	0,319	23,346	0,332
X_2 nivel de precios	−0,502	0,672	11,674	0,166
X_3 flexibilidad de precios	0,891	−0,186	33,362	0,477
X_4 imagen del fabricante	−0,041	0,190	0,961	0,009
X_5 servicio global	0,101	0,957	23,692	0,347
X_6 imagen del personal de ventas	0,071	0,224	0,126	0,027
X_7 calidad del producto	−0,222	0,163	9,293	0,026

de la discriminación. Esto puede hacerse sólo a través del examen de las cargas, pero se completa mediante una representación gráfica de las cargas discriminantes, como se describe en la siguiente sección.

Representación gráfica de las cargas discriminantes Para representar las diferencias en términos de variables predictoras, se pueden dibujar las cargas y los centroides de los grupos en un espacio discriminante reducido. Como se mencionó anteriormente, la representación más adecuada es mediante vectores de atributos extendidos y de los centroides de los grupos. La Tabla 5.24 muestra los cálculos extendidos tanto de las cargas discriminantes (empleadas por los vectores de atributos) como de los centroides de los grupos. El proceso de representación siempre contiene todas las variables incluidas en el modelo siguiendo el procedimiento por etapas (en nuestro ejemplo, las variables X_1, X_2 y X_3). Pero frecuentemente el investigador dibuja las variables que no están incluidas en la función discriminante si sus respectivos ratios F univariantes son significativos. Este procedimiento señala la importancia de variables colineales que no fueron incluidas en el modelo final por etapas. La importancia de incluir estas variables en el proceso de representación se demuestra en nuestro ejemplo. La variable X_7 se incluye en el modelo final, dado que es significativa en una parte de la varianza residual. La variable X_5 muestra diferencias mucho más fuertes entre los grupos pero no fue incluida debido a la colinealidad.

Los dibujos de los vectores de atributos extendidos para las cargas discriminantes rotadas se muestran en la Figura 5.13. Dibujando los vectores mediante este procedimiento, éstos apuntan a los grupos que tienen las mayores medias sobre la variable independiente respectiva y se alejan de los grupos que tienen las menores puntuaciones medias. Por ello, la interpretación del dibujo en la Figura 5.12 indica que la primera función discriminante (1) es la primera fuente de diferencia entre los grupos 1 y 2 frente al grupo 3. Además, la primera función se relaciona más estrechamente con la variable X_3 y algo con las variables X_1, X_2 y X_7 tienen cargas negativas. Esto se muestra no sólo en el gráfico, sino también a partir de un examen de las cargas discriminantes. Por ello, las características que diferencian a las empresas que como actividad primaria realizan decisiones de recompra directa (grupo 3) es que tienen percepciones más favorables sobre la rapidez en la entrega y la flexibilidad en el precio y percepciones menos favorables sobre el nivel de precios y la calidad del producto. Se pueden realizar perfiles similares para los dos grupos restantes de la primera función discriminante.

La segunda función discriminante (2) proporciona la distinción entre el grupo 1 frente a los grupos 2 y 3. La relación cercana entre los vectores X_2 y X_5 y la segunda función significa que las percepciones de la calidad del servicio y del producto describen mejor la segunda función discriminante.

En este momento, puede parecer que los gráficos son demasiado complicados. ¿Qué otro procedimiento es más simple a la vez que efectivo? El enfoque más fácil consiste en emplear las correlaciones rotadas (cargas) de la Tabla 5.18 y los centroides de los grupos. Las cargas a pie de página indican sobre qué función tiene mayor carga cada variable (por ejemplo, X_3 para la función 1 frente a X_2 y X_5 para la función 2). Para determinar entre qué grupos discrimina cada fun-

TABLA 5.24. Cálculo de los vectores de atributos extendidos y de las centroides de grupo en el espacio discriminante reducido

	Cargas de las funciones discriminantes rotadas		*Ratio F univariante*	*Coordenadas del espacio reducido*	
Variables independientes	*Función* 1	*Función* 2		*Función* 1	*Función* 2
X_1 rapidez del envío[a]	0,568	0,319	23,346	13,261	7,447
X_2 nivel de precios[a]	−0,502	0,672	11,674	−5,860	7,845
X_3 flexibilidad de precios[a]	0,891	−0,186	33,362	29,726	−6,205
X_4 imagen del fabricante	−0,041	0,190	0,961	−0,039[b]	0,183[b]
X_5 servicio global	0,101	0,957	23,692	2,393	22,673
X_6 imagen del personal de ventas	0,071	0,224	0,126	0,009[b]	0,028[b]
X_7 calidad del producto	−0,222	0,163	9,293	−2,063	1,515

	Centroides de grupo		*Valor F aproximado*		*Coordenadas del espacio reducido*	
Grupo	*Función* 1	*Función* 2	*Función* 1	*Función* 2	*Función* 1	*Función* 2
Grupo 1: Nueva tarea	−1,081	−1,167	55,148	15,162	−59,614	−17,694
Grupo 2: Recompra modificada	−0,952	0,879	55,148	15,162	−52,500	13,327
Grupo 3: Recompra directa	1,541	0,472	55,148	15,162	84,982	7,156

[a] Variables incluidas en la solución por etapas.
[b] Vectores no dibujados debido a un ratio F no significativo.

ción, obsérvense simplemente los centroides de los grupos y véase dónde se presentan las diferencias. Esto es meramente una valoración de la distancia, no una medida estadística, pero generalmente es suficiente. Por ejemplo, observando la función 1, vemos que la centroide para grupo 1 es de –1,081, para el del grupo 2 es de –0,952 y para el del grupo 3 de 1,541. A partir de esto podemos concluir que el origen primero de las diferencias para esta función es entre los grupos 1 y 2 frente al grupo 3. Un enfoque similar puede emplearse para la función 2. Sin embargo, como la función 1 representa sustancialmente más varianza que la función 2, se debe ser cauto al determinar la influencia de las variables basados en sus cargas sobre esta función.

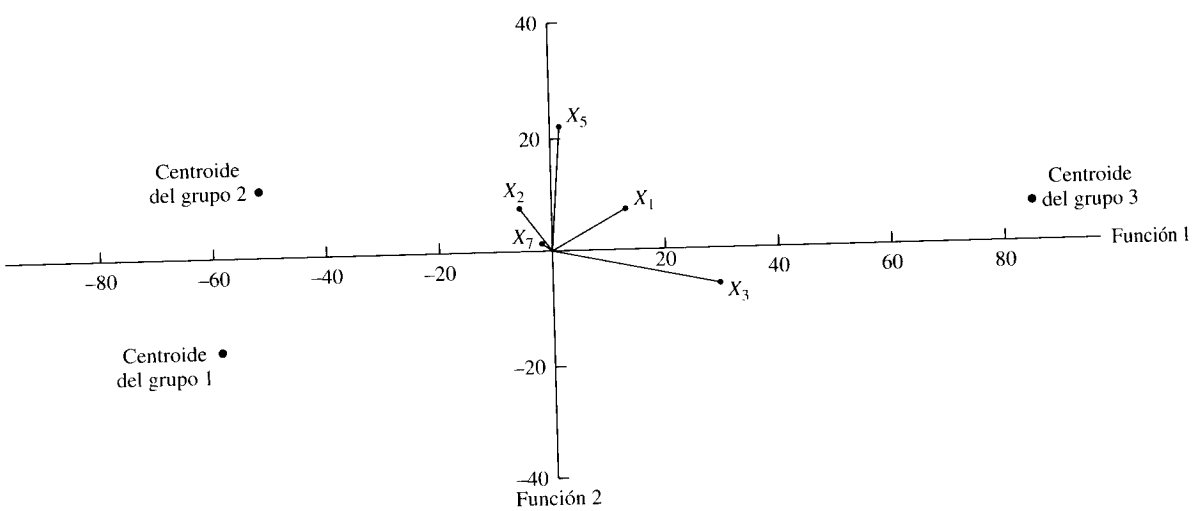


FIGURA 5.13. Dibujo de los vectores de atributos extendidos (variables) en el espacio discriminante reducido.

Sexto paso: Validación de los resultados discriminantes

La validez interna de los resultados discriminantes se apoya en los niveles de capacidad predictiva encontrados en la ampliación de la muestra. Para cada medida de la capacidad predictiva, las funciones discriminantes aplicadas a la ampliación de la muestra excedieron los valores críticos. La validación de los resultados discriminantes puede realizarse también mediante el método-*U*. Aquí, se estima el modelo discriminante dejando fuera un caso y prediciendo ese caso con el modelo estimado. Esto se hace para cada observación, de tal forma que cada observación nunca influye en el modelo discriminante que predice su clasificación de grupo. Los resultados de la clasificación cruzada se muestran en la Tabla 5.19. El ratio de aciertos global es mejor que el de los resultados para la muestra ampliada, y el bajo ratio para el grupo 2 mejora hasta el 60 por ciento, aunque el ratio de aciertos para el grupo 1 disminuye algo hasta el 76,2 por ciento. Con ratios de aciertos aceptables para la muestra ampliada y la validación cruzada, se confirma la validación del modelo discriminante.

Se recomienda también al investigador completar los procesos de validación utilizando perfiles de los grupos sobre conjuntos adicionales de variables y/o aplicar la función discriminante a otra(s) muestra(s) representativa(s) del total de la población o segmentos dentro de la población. Además el análisis de las clasificaciones erróneas ayudará a establecer qué variables adicionales se necesitan y/o si hay que revisar las clasificaciones de los grupos dependientes.

Una visión gerencial de conjunto

El análisis discriminante, orientado a la comprensión de las diferencias perceptivas de los clientes basadas en su tipo habitual de situación de compra con HATCO, es el origen de varios resultados. Primero, hay dos dimensiones de discriminación entre los tipos de situación de compra. La primera dimensión está tipificada por percepciones mayores de la rapidez en el envío y la flexibilidad de precios, que puede ser indicativa de cualidades relacionales. La segunda dimensión está mejor caracterizada por altas percepciones sobre el nivel de precios y el servicio, características más objetivas de la transacción.

Estos tres grupos pueden también perfilarse sobre estas dos dimensiones y las variables asociadas con cada dimensión para comprender las diferencias perceptivas entre ellas. El grupo de compra directa, que está caracterizado como una transacción rutinaria y repetitiva, tiene unas percepciones bastante más elevadas de HATCO que los otros dos grupos sobre la primera dimensión, quizá indicativo de una relación desde hace mucho tiempo. Los grupos 2 y 3, menores sobre la primera dimensión, se distinguen por la segunda dimensión, con el grupo de recompra modificada teniendo mayores percepciones de HATCO sobre el nivel de precios y el servicio cuando se compara con el grupo de nueva tarea. Estos patrones generales pueden extenderse al perfilar los grupos sobre variables independientes separadas y centrarse sobre variables diferenciadoras clave, en este caso X_1 (rapidez de envío), X_2 (nivel de precios), X_3 (flexibilidad de precios) y X_5 (servicio global). El análisis discriminante identifica las variables con más influencia para que la dirección de la empresa pueda desarrollar programas más concisos incorporando un pequeño conjunto de variables.

Los resultados también identifican un conjunto de observaciones de los grupos de nueva tarea y recompra modificada, que potencialmente representa bien un cuarto tipo de situación de compra o están caracterizadas por una variable todavía no incluida en el modelo. Tienen percepciones bastante precisas, pero no son habituales de los otros clientes en esos tipos de situaciones de compra. En muchos casos, son más parecidos al grupo de recompra directa en lo relativo a sus percepciones. Por tanto, a la dirección se le presenta un *input* para la gestión para la planificación estratégica y táctica a partir, no sólo de los resultados directos del análisis discriminante, también a partir de los errores en la clasificación.

Un ejemplo ilustrativo de regresión logística

Hay varios motivos por los que la regresión logística es una alternativa atractiva al análisis discriminante en tanto que la variable dependiente cuente sólo con dos categorías. En primer lugar, la regresión logística está menos influida que el análisis discriminante por las diferencias de varianzas-covarianzas entre los grupos, un supuesto básico del análisis discriminante. En segundo lugar, la regresión logística puede tratar con variables independientes categóricas fácilmente, mientras que en el análisis discriminante el uso de variables ficticias creaba problemas con las igualdades de varianzas-covarianzas. Finalmente, los resultados de la regresión logística son paralelos a los de la regresión múltiple en lo relativo a su interpretación y las medidas de diagnóstico caso a caso disponibles para el examen de los residuos. El siguiente ejemplo es idéntico al del análisis discriminante de dos grupos presentado previamente, utilizando esta vez la regresión logística para estimar el modelo. Como veremos, la regresión logística como alternativa al análisis discriminante cuenta con muchas ventajas, pero el investigador debe también examinar los resultados cuidadosamente para evitar el sobreajuste de los datos u otros problemas de estimación del modelo como los encontrados en este ejemplo.

Pasos primero, segundo y tercero: Diseño de objetivos, diseño de la investigación y supuestos estadísticos

Los temas tratados en estos tres primeros pasos del proceso de decisión son idénticos para el análisis discriminante de dos grupos y la regresión logística. El problema para el investigador consiste también en determinar si las diferencias en las percepciones de HATCO permiten distinguir entre clientes haciendo uso de la especificación de compra frente al análisis del valor total. La muestra de 100 clientes se divide en una muestra de análisis de 60 observaciones. Con las 40 observaciones restantes se forma la ampliación de la muestra o muestra de validación. Finalmente, dada la robustez del modelo de regresión logística respecto al incumplimiento del supuesto de igualdad de las matrices de varianzas-covarianzas entre grupos, la regresión logística resulta bastante adecuada en esta situación. Nos centramos ahora en los resultados que se derivan del uso de la regresión logística para estimar y comprender las diferencias entre estos dos tipos de clientes.

Cuarto paso: Estimación del modelo de regresión logística y valoración del ajuste global

Si volvemos a nuestra discusión sobre las diferencias de los grupos sobre las siete variables independientes (referido a la Tabla 5.5), recordamos que X_1, X_3 y X_7 cuentan con las mayores diferencias entre los dos grupos. En la regresión logística, una comparación análoga consiste en valorar la relación logística entre cada variable independiente y X_{11}, similar a las representaciones de las correlaciones utilizadas en la regresión. La Figura 5.14 muestra la relación logística para la siete variables independientes, junto con las correlaciones parciales. Como se vio examinando las medias, X_1, X_3 y X_7 tienen las mayores correlaciones y por tanto la mayor capacidad discriminante entre los grupos. En la relación logística de la Figura 5.13, la línea vertical dibujada en el 0,50 de valor predicho de la probabilidad es la puntuación de corte para hacer las predicciones de grupo. Por ejemplo, en el caso de X_1, las observaciones a la izquierda de la puntuación de corte se clasificarán en el grupo de especificación de compra ($X_{11} = 0$), mientras que las observaciones a la derecha de la puntuación de corte se clasificarán en el grupo de análisis del valor total ($X_{11} = 1$). Los casos de clasificación errónea pueden identificarse en cada grupo. Por ejemplo, para el grupo de especificación de compra al final de los gráficos, todas las observaciones a la derecha de la línea estarán mal clasificadas dentro del grupo de análisis del valor total. En algunos casos, como en X_4, la va-

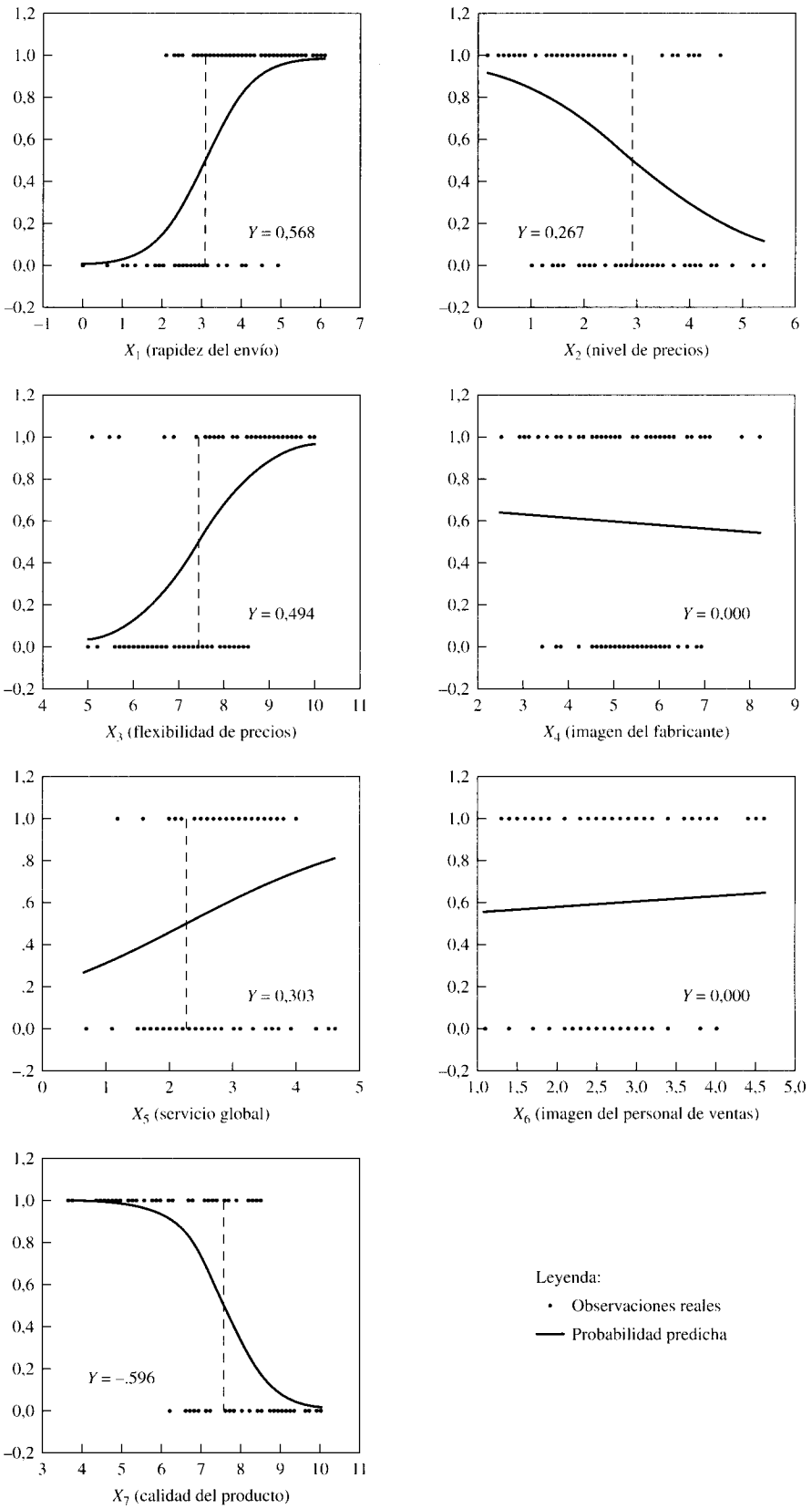


FIGURA 5.14. Relación logística entre X_{11} (especificación de compra) y las variables independientes.

riable independiente puede no hacer distinción entre los grupos y todos los casos se clasifican dentro del grupo mayor, en este caso el grupo 1. En estas situaciones, los valores de la probabilidad predicha nunca caen por debajo del 0,50.

Estimación del modelo

La regresión logística se estima de forma muy parecida a la regresión múltiple en la que se estima primero un modelo base para suministrar una referencia para la comparación (para más detalle, véase Capítulo 4). En la regresión múltiple, la media se emplea para establecer el modelo base y calcular la suma total de los cuadrados. En la regresión logística, se utiliza el mismo proceso, con la media utilizada en el modelo estimado, no para establecer la suma de los cuadrados, sino el valor del logaritmo de la verosimilitud. A partir de este modelo, se pueden hallar las correlaciones parciales para cada variable y elegir la variable más discriminante de acuerdo a los criterios de selección.

La Tabla 5.25 contiene los resultados del modelo base del análisis de la regresión logística. El valor del logaritmo de la verosimilitud (–2LL) aquí es de 78,859. El estadístico de puntuación, una medida de asociación utilizada en la regresión logística, junto con las correlaciones parciales para cada variable independiente, son indicadores de la variable seleccionada en el procedimiento por etapas. Se pueden utilizar varios criterios para orientar la incorporación: la mayor reducción en el valor –2LL, el mayor coeficiente de Wald, o la mayor probabilidad condicionada. En nuestro ejemplo, empleamos la reducción del criterio de razón del logaritmo de la verosimilitud.

Como se vio en la primera etapa del proceso de estimación, se seleccionó X_7 para su entrada (Tabla 5.26). Ésta se correspondía al mayor estadístico de puntuación y a la mayor correlación parcial. Aunque la entrada de X_7 en el modelo de regresión logística obtuvo un ajuste del modelo razonable, el examen de las variables independientes restantes que no estaban en la ecuación indicó que varias cumplían nuestro umbral de significación de 0,05 para su inclusión en el modelo y que, por tanto, se señalaba una posterior expansión del modelo. X_3, con el mayor estadístico de puntuación y correlación parcial, se seleccionó para la entrada en la etapa 2 (Tabla 5.27). Hubo una mejoría en todas las medidas de ajuste del modelo, desde un descenso del valor de –2LL a las diversas medidas R^2. De nuevo, aunque se obtuvieron niveles extremadamente altos de ajuste del modelo con dos variables, el examen de las variables no incluidas en la ecuación indicó que podían entrar otras variables en la solución por etapas. Sin embargo, cuando se introdujo X_5, la variable restante con la mayor correlación parcial y estadístico de puntuación, el modelo de regresión lo-

TABLA 5.25. Resultados del modelo base de la regresión logística

AJUSTE GLOBAL DEL MODELO
−2 log de la verosimilitud (−2LL): 78,859

VARIABLES NO INCLUIDAS EN LA ECUACIÓN

	Estadístico de puntuación	*Significación*	*Correlación parcial (r)*
X_1 rapidez del envío	27,476	0,000	0,568
X_2 nivel de precios	7,631	0,006	0,267
X_3 flexibilidad de precios	21,287	0,000	0,495
X_4 imagen del fabricante	0,840	0,360	0,000
X_5 servicio global	9,256	0,002	0,303
X_6 imagen del personal de ventas	1,631	0,202	0,000
X_7 calidad del producto	30,041	0,000	0,596

TABLA 5.26. Primer paso: Entrada de X_7 (calidad del producto) en el modelo de regresión logística por etapas

AJUSTE GLOBAL DEL MODELO

Medidas de bondad del ajuste	*Valor*	*Cambio en −2LL*		
			Valor	*Significación*
−2 log verosimilitud (−2LL)	37,524			
Bondad del ajuste	37,408	Desde el modelo base	41,335	0,000
«Pseudo» R^2	0,524	Desde el paso anterior	41,335	0,000
R^2 de Cox y Snell	0,498			
R^2 de Nagelkerke	0,681			

	Chi-cuadrado	*gl*	*Significación*
Hosmer y Lemeshow	2,664	8	0,9535

VARIABLES INCLUIDAS EN LA ECUACIÓN

Variable	*B*	*S.E.*	*Wald*	*Significación*	*r*	*Exp(B)*
X_7 calidad del producto	−1,896	0,495	14, 678	0,000	−0,401	0,150
Constante	14,581	3,794	14,774	0,000		

VARIABLES NO INCLUIDAS EN LA ECUACIÓN

	Estadístico de puntuación	*Significación*	*Correlación parcial (r)*
X_1 rapidez del envío	10,593	0,001	0,328
X_2 nivel de precios	0,214	0,643	0,000
X_3 flexibilidad de precios	15,614	0,000	0,415
X_4 imagen del fabricante	4,985	0,026	0,195
X_5 servicio global	6,669	0,010	0,243
X_6 imagen del personal de ventas	6,441	0,011	0,237

MATRIZ DE CLASIFICACIÓN

	Pertenencia al grupo predicho[a]					
	Muestra de análisis			*Ampliación de la muestra*		
Grupo real	*Grupo* 0	*Grupo* 1	*Total*	*Grupo* 0	*Grupo* 1	*Total*
Grupo 0: Especificación de la compra	17 (77,3)	5 (22,7)	22	14 (77,8)	4 (22,2)	18
Grupo 1: Análisis del valor total	4 (10,5)	34 (89,5)	38	3 (13,6)	19 (86,4)	22
Total	21	39	60	17	23	40

B = coeficiente logístico; S,E, = error estándar; Wald = estadístico de Wald; Signif, = nivel de significación; r = correlación; Exp(B) = coeficiente exponenciado.

[a] Los valores en paréntesis son los porcentajes correctamente clasificados (ratio de aciertos).

gística resultó inestable y dio lugar a coeficientes inapropiados e incluso indicó un ajuste perfecto en algunas medidas. Aún más, los coeficientes estimados tenían valores y niveles de significación inapropiados. Dada la naturaleza iterativa del proceso de estimación, el investigador siempre debería examinar los resultados para ver si ha tenido lugar este tipo de «sobreajuste» y seleccionar el modelo previo como el más apropiado. En este caso, seleccionamos el modelo de dos variables y lo utilizamos para una interpretación y valoración más extensas.

Significación estadística

Existen dos contrastes estadísticos para la significación del modelo final (véase Tabla 5.27). Primero, un contraste chi-cuadrado para el cambio en el valor del –2LL a partir del modelo base es

TABLA 5.27. Segundo paso: entrada de X_3 (flexibilidad de precios) en el modelo de regresión logística por etapas

AJUSTE GLOBAL DEL MODELO

Medidas de bondad del ajuste	*Valor*	*Cambio en −2LL*	*Valor*	*Significación*
−2 log verosimilitud (−2LL)	20,258			
Bondad del ajuste	58,967	Desde el modelo base	58,601	0,000
"Pseudo" R^2	0,743	Desde el paso anterior	17,266	0,000
R^2 de Cox y Snell	0,623			
R^2 de Nagelkerke	0,852			

	Chi-cuadrado	*gl*	*Significación*
Hosmer y Lemeshow	10,344	8	0,2417

VARIABLES INCLUIDAS EN LA ECUACIÓN

Variable	*B*	*S,E,*	*Wald*	*Significación*	*r*	*Exp(B)*
X_3 flexibilidad de precios	1,830	0,717	6,517	0,011	0,239	6,237
X_7 calidad del producto	−2,912	1,135	6,581	0,010	−0,241	0,054
Constante	8,329	5,110	2,657	0,103		

VARIABLES NO INCLUIDAS EN LA ECUACIÓN

	Estadístico de puntuación	*Significación*	*Correlación parcial (r)*
X_1 rapidez del envío	3,746	0,053	0,149
X_2 nivel de precios	3,641	0,056	0,144
X_4 imagen del fabricante	5,557	0,003	0,212
X_5 servicio global	8,824	0,003	0,294
X_6 imagen del personal de ventas	8,770	0,003	0,293

MATRIZ DE CLASIFICACIÓN

	Pertenencia al grupo predicho[a]					
	Muestra de análisis			*Ampliación de la muestra*		
Grupo real	*Grupo 0*	*Grupo 1*	*Total*	*Grupo 0*	*Grupo 1*	*Total*
Grupo 0: Especificación de la compra	21 (95,5)	1 (4,5)	22	15 (83,3)	3 (16,7)	18
Grupo 1: Análisis del valor total	0 (0,0)	38 (100,0)	38	2 (9,1)	20 (90,9)	22
Total	21	39	60	17	23	40

B = coeficiente logístico; S,E, = error estándar; Wald = estadístico de Wald; Signif, = nivel de significación; r = correlación; Exp(B) = coeficiente exponenciado.

[a] Los valores en paréntesis son los porcentajes correctamente clasificados (ratio de aciertos).

comparable al contraste global F de la regresión múltiple. En el modelo de dos variables, esta reducción era estadísticamente significativa a un nivel del 0,000. Además, la medida de Hosmer y Lemeshow de ajuste global [10] cuenta con un contraste estadístico que indica que no había una diferencia estadísticamente significativa entre las clasificaciones observadas y predichas. Estas dos medidas, en combinación, suministran apoyo para aceptar el modelo de dos variables como un modelo de regresión logística significativo y adecuado para un examen posterior.

Los coeficientes estimados para las dos variables independientes y la constante pueden también evaluarse para la significación estadística. El estadístico de Wald se utiliza para valorar la significación, excepto en casos en que el coeficiente es extremadamente elevado, en que se utiliza el estadístico de puntuación. Ambos coeficientes son estadísticamente significativos a un nivel del 0,01, aunque la constante es significativa sólo a un nivel de 0,10. Por tanto, las variables individuales son significativas y deberían ser interpretadas.

Valoración del ajuste global del modelo

Para valorar el ajuste global existen varias medidas. Primero, el valor –2LL. En el modelo de una variable (véase Tabla 5.26), el valor –2LL se reduce sobre el modelo base de 78,859 a 37,524, un decrecimiento de 41,335. Menores valores de la medida –2LL indican un mejor ajuste del modelo. La bondad de la medida de ajuste compara las probabilidades predichas con las probabilidades observadas con mayores valores indicando un mejor ajuste. No existe límite inferior o superior para esta medida y el valor para el modelo de variable única es de 37,408. A continuación, existen tres medidas comparables a la medida R^2 de la regresión múltiple. La R^2 de Cox y Snell opera de la misma forma, con los mayores valores, indicando un mejor ajuste del modelo. Sin embargo, esta medida está limitada en que no puede alcanzar el máximo valor de 1, por lo que Nagelkerke propuso una modificación que tenía el rango de 0 a 1. En nuestro caso, el valor de Cox y Snelles es de 0,498 y el valor de Nagelkerke es de 0,681. La tercera medida «pseudo» R^2 está basada en la mejora en el valor –2LL. El valor de 0,524 para el modelo de variable única se calcula como:

$$R^2_{\text{logit}} = \frac{0{,}2\text{LL}_{\text{nulo}} - (-2\text{LL}_{\text{modelo}})}{-2\text{LL}_{\text{nulo}}}$$

$$= \frac{78{,}859 - 37{,}524}{78{,}859}$$

$$= \frac{41{,}335}{78{,}859} = 0{,}524$$

La medida final de ajuste del modelo es el valor de Hosmer y Lameshow, que mide la correspondencia de los valores reales y predichos de la variable dependiente. En este caso, el mejor ajuste del modelo viene indicado por una menor diferencia entre la clasificación observada y la predicha. Un valor chi-cuadrado no significativo indica un buen ajuste del modelo.

En el modelo de dos variables, mejoran todas las medidas de ajuste del modelo. El valor –2LL decreció hasta 20,258. El valor R^2 pasó de 0,623 a 0,852, los dos mejor respecto del modelo de una variable e indicativo de un buen ajuste del modelo cuando se comparó con los valores R^2 normalmente encontrados en la regresión múltiple. La medida de Hosmer y Lebeshow todavía muestra no significatividad, indicando que no hay diferencia en la distribución de los valores dependientes reales y predichos.

Finalmente, las matrices de clasificación, idénticas en naturaleza a las utilizadas en el análisis discriminante, muestran ratios de aciertos extremadamente elevados y de casos correctamente clasificados para el modelo de dos variables. Los ratios de aciertos globales son del 98,3 por ciento y del 87,5 por ciento para las muestras de análisis y ampliada, respectivamente. De la misma forma, los ratios de aciertos de grupo individuales son consistentemente elevados y no son indicativos de problemas al predecir cualquiera de los dos grupos. El modelo de dos variables, incluyendo X_3 y X_7, demuestra un ajuste del modelo excelente y significación estadística a un nivel de modelo global además de para las variables incluidas en el modelo.

Diagnóstico caso a caso

En análisis de las clasificaciones erróneas de las observaciones individuales puede suministrar una mayor información sobre posibles mejoras del modelo, pero en este caso hay un total de sólo seis clasificaciones erróneas en las muestras de análisis y ampliada juntas, proporcionando un fundamento inadecuado para realizar un posterior análisis. El diagnóstico caso a caso está disponible al igual que los residuos y las medidas de influencia, pero se usan poco en esta situación (para una discusión más detallada de medidas como la distancia de Cook y la DFBETA, véase Apéndice del Capítulo 4). Además, dados unos bajos niveles de clasificación errónea, no se lleva a cabo un análisis posterior de las mismas.

Quinto paso: Interpretación de los resultados

El valor teórico del modelo de regresión logística es bastante parecido al del análisis discriminante de dos grupos, aunque con una variable independiente menos. En el caso del análisis discriminante, X_1, X_3 y X_7 se incluyeron en la solución por etapas. Pero la regresión logística incluyó sólo a X_3 y X_7 (véase Tabla 5.27). Las consecuencias para ambos análisis eran similares: la flexibilidad de precios (X_3) está asociada positivamente con la variable dependiente y la calidad del producto (X_7) negativamente. Dado que la variable dependiente (X_{11}) tiene dos grupos —especificación de compra ($X_{11} = 0$) y análisis del valor total ($X_{11} = 1$)— los coeficientes implican que las empresas que usan el análisis del valor total tienen menores percepciones de la calidad del producto mientras que cuentan con mayores percepciones de la flexibilidad de precios. Debido a que tanto el análisis discriminante como la regresión logística predicen pertenencia a un grupo, los coeficientes relacionan la media relativa de cada grupo con las variables independientes.

Sexto paso: Validación de los resultados

La validación del modelo de regresión logística se realiza en este ejemplo de la misma forma que se hizo en el análisis discriminante, mediante una muestra de análisis y una ampliación de la muestra. Aunque se aprecie una mejora en el ratio de aciertos en la muestra de análisis con la regresión logística, las ampliaciones de la muestra son prácticamente idénticas para el análisis discriminante y para la regresión logística. Esto lleva a la conclusión de que ambos métodos cuentan con un apoyo empírico importante y su validación en muestras separadas está aproximadamente al mismo nivel.

Una visión gerencial de conjunto

La regresión logística supone una alternativa respecto al análisis discriminante que puede resultar más «cómoda» a muchos investigadores debido a sus parecidos con la regresión múltiple. Dada su robustez frente a condiciones en los datos que pueden afectar negativamente al análisis discriminante (por ejemplo, matrices desiguales de varianzas-covarianzas), la regresión logística resulta la técnica de estimación preferida en muchas aplicaciones. En este ejemplo, la regresión logística suministró un ligero aumento en la exactitud predictiva con un valor teórico más simple que contaba básicamente con la misma interpretación, sólo que con una variable menos. A partir de los resultados de la regresión logística, el investigador puede centrarse en la relación de intercambio de la flexibilidad de precios frente a la calidad del producto cuando le preocupa exclusivamente la predicción. Sin embargo, para aumentar el grado de comprensión, las variables no incluidas en el análisis pero que muestran diferencias entre los grupos deberían completar estos resultados. La colinealidad entre las variables puede hacer que resulte redundante la capacidad discriminatoria entre variables, pero la redundancia no hace irrelevantes a las variables desde el punto de vista de la explicación.

Resumen

Se ha presentado la naturaleza, conceptos y enfoque subyacente del análisis discriminante múltiple y de la regresión logística. Se incluyeron las líneas básicas de su aplicación e interpretación para aclarar aún más los conceptos metodológicos. Se presentaron ejemplos ilustrativos de soluciones tanto para dos como para tres grupos utilizando la base de datos de HATCO. Estas aplicaciones trataron las principales cuestiones que surgen para familiarizarse con la aplicación del análisis discriminante y la regresión logística y para seleccionar entre ambos métodos en diferentes situaciones.

El análisis discriminante múltiple ayuda a comprender y explicar los problemas de investigación que incluyen una única variable dependiente categórica y varias variables independientes métricas. También es posible que las variables independientes estén compuestas de conjuntos de datos mezclados (tanto métricos como no métricos) si las variables no métricas se codificar como variables ficticias (0 – 1). El resultado del análisis discriminante puede ayudar a perfilar las características de los sujetos dentro de los grupos y a asignarlos a sus grupos correctos. Las aplicaciones potenciales del análisis discriminante tanto en problemas empresariales como no empresariales son numerosas.

Algunos de los conceptos presentados en este capítulo están basados en temas tratados en los Capítulos 3, 4 y 6. Por ello, se recomienda que estos tres capítulos se estudien conjuntamente.

Cuestiones de revisión

1. ¿Cómo diferenciaría entre análisis discriminante múltiple, análisis de regresión y análisis de la varianza?
2. ¿Cuándo emplearía la regresión logística en lugar del análisis discriminante? ¿Cuáles con las ventajas y desventajas de esta decisión?
3. ¿Qué criterio podría utilizar para decidir si detenerse o no en un análisis discriminante después de estimar la(s) función(es) discriminante(s)?
4. ¿Qué procedimiento llevaría a cabo para dividir su muestra en grupos de análisis y ampliación? ¿Cómo cambiaría este procedimiento si su muestra constase de menos de 100 individuos u objetos?
5. ¿Cómo determinaría la puntuación de corte óptima?
6. ¿Cómo determinaría si es suficientemente alta la capacidad clasificatoria de la función discriminante en relación con una clasificación puramente aleatoria?
7. ¿En qué se diferencia un análisis discriminante de dos grupos de un análisis de tres grupos?
8. ¿Por qué un investigador extendería los datos de los centroides y de las cargas para representar gráficamente una solución del análisis discriminante?
9. ¿Cómo tratan la regresión logística y el análisis discriminante la relación entre las variables independientes y dependiente?
10. ¿Cuáles son las diferencias en la estimación y en la interpretación entre la regresión logística y el análisis discriminante?

Bibliografía

1. Cohen, J. (1977), *Statistical Power Analysis for the Behavioral Sciences*, rev. ed. New York: Academic Press.
2. Crask, M., and W. Perreault (1977), «Validation of Discriminant Analysis in Marketing Research.» *Journal of Marketing Research* 14 (February): 60–68.
3. Dillon, W. R., and M. Goldstein (1984), *Multivariate Analysis: Methods and Applications.* New York: Wiley.
4. Frank, R. E., W. E. Massey, and D. G. Morrison (1965), «Bias in Multiple Discriminant Analysis.» *Journal of Marketing Research* 2(3):250–58.
5. Gessner, Guy, N. K. Maholtra, W. A. Kamakura, and M. E. Zmijewski (1988), «Estimating Models with Binary Dependent Variables: Some Theoretical and Empirical Observations.» *Journal of Business Research* 16(1):49–65.
6. Green, P. E., D. Tull, and G. Albaum (1988), *Research for Marketing Decisions*. Upper Saddle River, N.J.: Prentice Hall.
7. Green, P. E. (1978), *Analyzing Multivariate Data.* Hinsdale, Ill.: Holt, Rinehart, and Winston.
8. Green, P. E., and J. D. Carroll (1978), *Mathematical Tools for Applied Multivariate Analysis.* New York: Academic Press.
9. Harris, R. J. (1975), *A Primer of Multivariate Statistics.* New York: Academic Press.
10. Hosmer, D. W., and S. Lemeshow (1989), *Applied Logistic Regression.* New York: Wiley.

11. Huberty, C. J. (1984), «Issues in the Use and Interpretation of Discriminant Analysis.» *Psychological Bulletin* 95: 156–71.
12. Huberty, C. J., J. W. Wisenbaker, and J. C. Smith (1987), «Assessing Predictive Accuracy in Discriminant Analysis.» *Multivariate Behavioral Research* 22 (July): 307–29.
13. Johnson, N., and D. Wichern (1982), *Applied Multivariate Statistical Analysis.* Upper Saddle River, N.J.: Prentice Hall.
14. Morrison, D. G. (1969), «On the Interpretation of Discriminant Analysis.» *Journal of Marketing Research* 6(2): 156–63.
15. Perreault, W. D., D. N. Behrman, and G. M. Armstrong (1979), «Alternative Approaches for Interpretation of Multiple Discriminant Analysis in Marketing Research.» *Journal of Business Research* 7: 151–73.

LECTURA COMPLEMENTARIA

Factores determinantes del nivel de equipamiento doméstico. Un contraste empírico

CARLOS FLAVIÁN BLANCO
EVA MARTÍNEZ SALINAS
YOLANDA POLO REDONDO
Universidad de Zaragoza

ESIC, n.º 85
Julio-septiembre, 1994

1. INTRODUCCIÓN

En los últimos años, se ha obsevado un importante incremento en el equipamiento doméstico en los hogares de nuestro país. Productos como microondas, vitrocerámica, lavavajillas, etc., están presentes cada vez más en las viviendas de muchas familias. Resulta, por tanto, de interés conocer qué factores han condicionado la adquisición de determinados productos que comportan un mayor bienestar en el hogar. Nuestro objetivo en este trabajo es contrastar si determinadas características socioeconómicas, tales como: nivel de ingresos, ciclo de vida familiar, número de hijos, nivel de formación del cabeza de familia... explican una mayor comodidad a la hora de realizar las tareas domésticas. Junto a ellos hemos considerado la incorporación de la mujer al mundo del trabajo que, desde nuestro punto de vista, puede resultar un factor decisivo a la hora de explicar un hogar mejor equipado. Este último factor supone importantes cambios a la hora de realizar las compras por parte de la unidad familiar. Supone, por un lado, una mayor demanda de estos productos y, por otro, un menor tiempo libre.

Los hogares van a ser unidades de consumo donde sus miembros van a intentar que las compras de bienes y la utilización de servicios les proporcionen una mayor satisfacción, lo cual a su vez se traducirá en un importante ahorro de tiempo pero a su vez, algunos de estos bienes por las características particulares de los mismos, mantenimiento y utilización, requerirán un consumo de tiempo. Goldberg (1977), Linder (1970) y Schary (1971) afirman que conforme aumenta la adquisición de productos en el hogar que necesitan mantenimiento, los consumidores perciben que disponen de menos tiempo. Frente a estas afirmaciones, otros autores (McCall, 1977; Rosenberg y Hirschman, 1980; Berkowitz, Walton y Walter, 1979) consideran que para solucionar la falta de tiempo en los hogares se hace uso de determinados productos y servicios que ahorran tiempo.

Por esta razón, a la hora de analizar los productos adquiridos por los hogares, podríamos hablar de productos que ahorran tiempo, bien porque realizan alguna tarea que antes tenía que realizar una persona, o bien porque su mantenimiento requiere menos tiempo que el producto anterior al que sustituye; y de productos que consumen tiempo, bien porque son productos que se utilizan para el ocio, o bien porque disponer de ellos supone realizar alguna tarea que no se haría en caso de no tenerlos.

Parecería lógico pensar que la disponibilidad de tiempo de los individuos que componen la unidad familiar lleve o no a la adquisición de determinados productos y a la utilización de determinados servicios. El hecho de que la mujer trabaje o no fuera del hogar, así como el número de horas que lo hace, determinarán el tiempo que puede dedicar a las tareas domésticas y por tanto a la manera en que lo hace (Strober y Weinberg, 1980). Shaninger y Allen (1981) proponen una clasificación del status de la mujer como determinante de la adquisición de determinados productos. Los ingresos y el ciclo de vida familiar se consideran buenos predictores del comportamiento del consumidor en este aspecto, incluso mejores que el hecho de que la mujer trabaje o no fuera del hogar (Douglas, 1976; Feldman y Hornik, 1981; Reilly, 1982).

El propósito principal del trabajo va a ser analizar si existe o no relación entre el nivel de equipamiento de los hogares y determinados condicionantes de tipo socioeconómico. Se usarán para el estudio productos del consumo duradero que facilitan las tareas del hogar (lavavajillas y vitrocerámica) o que por su naturaleza están asociados a familias con menor disponibilidad de tiempo (frigorífico «combi» y microondas). Como determinantes de carácter socioeconómico, se han considerado los ingresos, el tamaño de la unidad familiar, nivel de estudios de hombre y mujer y el status de trabajo de ambos; también se incluye el número de automóviles como variable que refleja el poder de compra del hogar.

Nuestro análisis se desarrollará en dos partes. En una primera se tratará de contrastar empíricamente si la renta recoge o no toda la información sobre el nivel socioeconómico de la población estudiada, y una segunda diferenciando las mujeres que trabajan fuera del hogar de las que no lo hacen, con el fin de conocer si existen diferencias entre ambos grupos a la hora de adquirir determinados productos.

Antes de llevar a cabo el análisis empírico, se comentan algunos de los trabajos más relevantes que han estudiado este tema, y que han servido de base para este análisis.

2. REVISIÓN DE LITERATURA

A lo largo de la literatura, se ha estudiado cómo influyen determinadas variables en la adquisición de determinados productos y en la utilización de algunos servicios, pero la mayoría de ellos han puesto énfasis en analizar la influencia del hecho de que la mujer trabaje o no fuera del hogar.

Se encuentran teorías contradictorias respecto a la relación entre adquirir productos de consumo duradero y trabajar o no fuera del hogar. Así Galbraith (1973), plantea que para hogares con el mismo nivel de ingresos, las familias cuya mujer trabaja fuera del hogar gastan menos en productos de consumo duradero que las familias en que la mujer no trabaja. Esta teoría se basa en considerar que las mujeres que trabajan disponen de menos tiempo para la utilización y el mantenimiento de estos productos. En contrapartida, Mincer (1960), propone la idea contraria, las mujeres que trabajan fuera del hogar gastan más en estos productos; esta idea considera que el trabajo de la mujer es transitorio y que por tanto al percibir más ingresos durante un período de tiempo, éstos se utilizan para el consumo en lugar de para el ahorro. Drucker (1976), partiendo de la misma idea, concluye que estos ingresos «extraordinarios» se utilizan en compras «extraordinarias». Relaciones positivas entre el trabajo de la mujer fuera del hogar y el consumo de determinados productos de consumo duradero, para familias con el mismo nivel de ingresos, las encontramos en Mallan (1968) y Agarwala y Drinkwater (1972).

La teoría económica parece indicar que familias con el mismo nivel de ingresos pero que difieren entre sí en el hecho de que en unas, la mujer trabaja fuera del hogar y en otras no, adquirirán distintos productos y servicios, buscando las primeras aquellos que les proporcionan ahorro de tiempo. Strobeg and Weinberg (1977), Weinberg y Winer (1983), analizan esta afirmación. En el primer trabajo se realiza un análisis discriminante para ver qué variables discriminan más a la hora de poseer o no determinados productos de consumo duradero y utilizar o no algunos servicios, llegando a la conclusión de que el hecho de que la mujer trabaje o no fuera del hogar no va a discriminar entre los dos tipos de familias, teniendo importancia por otro lado variables como la movilidad geográfica, los ingresos y el ciclo de vida familiar. Weinbeg y Winer corroboran esta afirmación repitiendo el estudio con la misma fuente de información varios años después. Ambos trabajos analizan también la relación entre la cantidad de gasto en productos y servicios respecto a las mismas variables, llegando a la conclusión de que la variable trabajar fuera del hogar no es relevante, siendo más importante los ingresos.

Durante los años ochenta encontramos una serie de investigaciones que tratan de concretar más, intentando analizar la relación entre el status social de la familia y el consumo de productos y servicios de conveniencia que ayudan en el trabajo del hogar o lo sustituyen (Andersom, Douglas, Reynols, Crask y Well...). En ninguno de estos trabajos llegan a encontrarse relaciones positivas entre ambos aspectos. Reilly (1982), afirma que los ingresos están relacionados con la adquisición de productos de consumo duradero, y éstos van a estar correlacionados con el status. Por tanto, aunque no se encuentran relaciones estadísticamente significativas entre el status familiar y la adquisición de productos que ahorran tiempo, se puede observar de manera indirecta.

Las mujeres que trabajan fuera de casa, además de aumentar el equipamiento del hogar con productos que ahorren tiempo, pueden adoptar distintas estrategias a la hora de realizar las tareas del hogar. Stroberg y Weinberg (1980), proponen cinco estrategias:

- Sustituir la realización de las tareas del hogar, manteniendo la calidad y la frecuencia de las mismas, mediante un alto nivel de equipamiento en el hogar.
- Sustituir la realización de las tareas del hogar, manteniendo la calidad y la frecuencia de las mismas, mediante la ayuda de otras personas (marido, hijos o servicio doméstico).
- Reducir la calidad y la frecuencia de realización de las tareas y/o llevarlas a cabo de manera más intensiva y eficiente.
- Disminuir, si es el caso, el tiempo dedicado a actividades voluntarias.
- Disminuir el tiempo libre y/o el dedicado a descansar.

En este estudio se demuestra, como en anteriores, que no existen diferencias entre el equipamiento del hogar de las mujeres que trabajan fuera del hogar de las que no lo hacen, pero sí se observan en los aspectos referidos al resto de estrategias.

Nickols y Fox (1983), llevan a cabo una réplica más extensa del trabajo anterior, llegando a las mismas conclusiones: variables como los ingresos, la edad del hijo más joven y el área de residencia van a influir en el tiempo dedicado a las tareas del hogar y en la ayuda, contratada a otros miem-

bros de la unidad familiar, que utilizan las mujeres que trabajan fuera del hogar.

Lovingood y McCullough (1986), afirman que casi no existen diferencias en el tiempo dedicado a las tareas del hogar entre los hogares que disponen de determinados productos del hogar y los que no los tienen; van a ser factores económicos y características familiares las que van a influir en la adquisición de los mismos.

En los últimos años, encontramos algunos trabajos que continúan profundizando en este tema. Así, Bryant (1988), critica los estudios anteriores afirmando que existe una relación entre el trabajo de la mujer y el gasto realizado en productos de consumo duradero, indicando que éstos no son sustitutivos del tiempo que la mujer dedica al hogar, sino complementarios.

Kim (1989), se centra en analizar la adquisición de productos de consumo duradero en el hogar por parte de las mujeres trabajadoras, así como la utilización de productos de alimentación de fácil preparación. La conclusión más relevante de este estudio es que se encuentran diferencias en la adquisición de algunos de los productos que ahorran tiempo, entre las familias en que la mujer trabaja fuera del hogar y en las que no lo hace, concluyéndose lo mismo en cuanto a la utilización de productos de alimentación de fácil preparación.

Para terminar, Oropesa (1993) analiza la adquisición de un producto considerado como ahorrador de tiempo en los hogares, el microondas, viendo el efecto de diferentes variables, entre ellas la participación de la mujer en la fuerza de trabajo, desde sus primeros años de aparición en el mercado (período 1978-1989).

El Cuadro 5.L.1 resume los análisis realizados y las conclusiones más relevantes de los trabajos comentados anteriormente.

3. METODOLOGÍA DEL ESTUDIO DE CAMPO

El universo objeto de estudio se limitó a Aragón y el ámbito fueron las familias residentes en el área metropolitana de Zaragoza. La recogida de información se realizó entre los días 11 y 30 de abril de 1994.

Para la selección de los individuos de la muestra, se realizó una estratificación del universo por áreas de residencia, realizando una afijación proporcional del número de encuestas a cada una de ellas.

La encuesta se realizó de forma personalizada mediante visita al domicilio del encuestado, que previamente había sido seleccionado a través del método de rutas aleatorias (Random Routes) con puntos de partida previamente fijados.

Se realizaron un total de 400 encuestas que posteriormente fueron sometidas a análisis para contrastar su validez. En este proceso fueron eliminadas 38 de ellas al encontrar indicios de contradicción en las respuestas del encuestado o por carecer de partes esenciales para el análisis. Tras el depurado de la información, pudo contarse un total de 362 encuestas válidas para el estudio.

De acuerdo con las expresiones matemáticas apropiadas, considerando un nivel de confianza del 95,5 por ciento, el error máximo aceptado es el de 65,25 por ciento.

4. ANÁLISIS Y RESULTADOS

La técnica de análisis utilizada, en primer lugar, ha sido el discriminante. Esta técnica clasificatoria supone incluir a los individuos que han respondido la encuesta en dos categorías: adoptante o no de cada tipo de producto.

El análisis discriminante analiza la asociación que existe entre la pertenencia a un grupo y las características descriptivas del individuo clasificado en ese grupo. Para llevar a cabo el análisis discriminante, se ha utilizado el algoritmo de estimación de la función discriminante contenido en el programa BMDP-7M en su versión de incorporación sucesiva, stepwise, de variables explicativas.

En este programa las variables son añadidas secuencialmente, es decir, el programa va introduciendo en cada etapa aquella variable explicativa que posee el mayor valor de «F entrada/salida». El proceso es continuo y determina la importancia de las variables por su orden de entrada.

Los resultados de este análisis se presentan en los Cuadros 5.L.2, 5.L.3 y 5.L.4, cada uno de ellos recoge todas las variables dependientes para distintos análisis efectuados. El Cuadro 5.L.2 corresponde al análisis discriminante utilizando a todos los encuestados los Cuadros 5.L.3 y 5.L.4, corresponden a la muestra de mujeres que trabajan fuera del hogar y a las que no lo hacen, respectivamente.

Dichos cuadros incluyen información sobre las estadísticas de las variables explicativas (1) en la fase en que son incorporadas al modelo y el valor de su estadístico F para cada producto; el número de personas que han adquirido cada producto y el porcentaje de individuos correctamente clasificados.

Globalmente, podemos concluir que los resultados estadísticos de este análisis son satisfactorios, no pudiendo rechazarse la hipótesis general de que existe una relación entre la adopción de determinadas características socioeconómicas. Esta afirmación estaría basada en la presencia de variables estadísticamente significativas en la mayoría de las funciones de clasificación estimadas, así como en los porcentajes de correcta clasificación superiores al 63 por ciento que se obtienen para los distintos procedimientos de reclasificación.

El Cuadro 5.L.2, presenta los resultados obtenidos al aplicar el discriminante a los datos. La variable dependiente toma el valor 1 cuando la familia encuestada dispone de ese producto, y 0 en caso contrario.

Como variables explicativas hemos incluido la renta (en unidades monetarias), el número de automóviles (como medida del grado de equipamiento de las familias (medido por

(1) REN = renta, NAU = número de automóviles, TUF = tamaño unidad familiar, TH = status de trabajo del hombre, TM = status de trabajo de la mujer, EH = Nivel de estudios del hombre, EM = nivel de estudios de la mujer.

CUADRO 5.L.1. Resumen de las investigaciones mas relevantes sobre los condicionantes de compra de distintos productos de consumo duradero

Autor año	*Tipo de muestra*	*Productos considerados*	*Variables utilizadas*	*Técnicas de análisis*	*Resultados más relevantes*
Strober y Weinberg 1977.	Panel de consumidores.	Secadora, lavavajillas, horno, frigorífico.	Ingresos, movilidad geográfica, ciclo de vida familiar, trabajo de la mujer, etc.	Análisis discriminante. Análisis de regresión.	Ingresos y movilidad son las variables que más discriminan. El trabajo de la mujer no es un factor relevante.
Strober y Weinberg 1980.	Panel de consumidores.	Lavavajillas, microondas, secadora, congelador, horno, lavadora, frigorífico.	Ingresos, ciclo de vida familiar, trabajo de la mujer fuera del hogar.	Análisis discriminante.	El trabajo de la mujer no es un factor discriminante.
Reilly 1982.	Muestra representativa.	Productos de consumo duradero sin especificar.	Ingresos mujer y familiar, educación mujer, status de trabajo, etc.	Modelo de ecuación estructural desarrollado en el trabajo.	Los ingresos son la variable que más discrimina. El status influye ya que está correlacionado con los ingresos.
Weinger y Winer 1983.	Panel de consumidores.	Lavavajillas, secadora, frigorífico, lavadora, horno.	Ingresos, movilidad geográfica, ciclo de vida familiar, trabajo de la mujer.	Análisis discriminante. Análisis Tobit.	Ingresos, movilidad y ciclo de vida son las más relevantes. El trabajo de la mujer no es un factor determinante.
Nickols y Fox 1983.	Muestra estratificada representativa.	Microondas, lavavajillas, horno, autolimpiable, secadora, triturador de basura, frigorífico-congelador.	Ingresos, edad del hijo menor, trabajo de la mujer fuera del hogar, status de trabajo del marido y de la mujer, área de residencia.	Análisis discriminante.	La variable que más discrimina son los ingresos. El trabajo de la mujer y su status no forman parte de la función discriminante.
Lovingood y McCullough 1986.	Muestra representativa.	Aspiradora, secadora, horno autolimpiable, lavadora, frigorífico, triturador de basura, microondas.	Ingresos, ocupación del encargado del hogar, horas de trabajo, ocupación de cónyuge, edad del hijo más joven, área de residencia.	Análisis de la chi-cuadrado.	En la adquisición de electrodomésticos van a influir factores económicos y características familiares, más que el ahorro de tiempo que algunos suponen.
Bryant 1988.	Muestra representativa.	Productos de consumo duradero sin especificar.	Edad, educación, tamaño familiar, edad hijo más joven e hijo mayor, ingresos del marido, área de residencia, raza.	Modelo desarrollado en el trabajo. Análiis M.C.O., Probit y Tobit.	Existe relación entre el trabajo de la mujer y el gasto en productos de consumo duradero.
Kin 1989.	Muestra representativa.	Lavadora, secadora, frigorífico, congelador, lavavajillas, aspiradora, microondas.	Status del trabajo de la mujer, propiedad del hogar, ciclo de vida familiar, cultura, tamaño familiar, ingresos.	Análisis discriminante.	El trabajo de la mujer va a influir en la adquisición de productos que ahorran tiempo.
Oropesa 1993.	Muestra estratificada representativa.	Microondas.	Participación de la mujer en la fuerza de trabajo, edad del hijo más joven, edad de la mujer, región de residencia, tamaño familiar, año de adquisición.	Análisis de regresiones logísticas.	El status de trabajo de la mujer va a ser un importante predictor en la adquisición de productos que ahorran tiempo.

CUADRO 5.L.2. Resultados del análisis discriminante para diferenciar entre familias compradoras y no compradoras

Variable	*REN*	*NAU*	*TUF*	*TH*	*TM*	*EH*	*EM*	*Número compradores*	*Porcentaje clasificados correctamente*
Frigorífico	1.º 10,6	—	—	—	—	—	—	316	68
Microondas	1.º 21,9	—	—	—	—	2.º 14,3	—	142	63
Lavavajillas	1.º 103,6	—	—	—	2.º 8,2	—	3.º 4,37	169	76
Vitrocerámica	1.º 29,4	—	—	—	—	—	—	134	64
Automóvil	2.º 25,3	*	—	—	1.º 38,3	—	—	301	73

* Variable no incluida.

el número de miembros), el status de trabajo del hombre y la mujer de la familia (medido a partir de una serie de categorías) (2) y el nivel de estudios del cabeza de familia medido de la misma manera (3).

De los cinco productos de consumo duradero analizados tan sólo se encuentra como variable explicativa de las diferentes adopciones la renta que entra siempre en primer lugar. Tanto el frigorífico «combi», como la cocina vitrocerámica sólo pueden explicar su compra en familias con un determinado nivel de ingresos. Estos resultados están de acuerdo con los encontrados en la literatura. Stroberg y Weinberg (1977, 1988) y Weinberg y Winer (1983), utilizan ingresos, ciclo de vida familiar, movilidad geográfica y trabajo de la mujer, como variables explicativas de la adopción, concluyendo que son los ingresos la variable que más discrimina. Nickols y Fox (1983) llegan a la misma conclusión, seleccionando a la variable ingresos como más discriminante de entre otras muchas.

Aparece un conjunto de dos productos, microondas y automóviles explicados por dos variables. En el primero, la variable más importante son los ingresos, seguida del grado de formación del cabeza de familia. En cuanto al automóvil, la variable más discriminante es el puesto de trabajo desarrollado por el hombre, seguida del nivel de renta. En este último caso se ha excluido del análisis la variable «número de automóviles» por la fuerte correlación con la variable dependiente.

El producto que presenta unos mejores resultados es el lavavajillas. La variable que entra en primer lugar son los ingresos, seguida del nivel de formación y del trabajo de la mujer. Este producto es el que consigue clasificar correctamente un porcentaje más alto (76 por ciento).

Los Cuadros 5.L.3 y 5.L.4 resumen los resultados de aplicar un análisis discriminante a la muestra distribuida en dos grupos: mujeres que trabajan fuera del hogar frente al resto de ellas.

Para el caso de mujeres amas de casa, en paro y pensionistas, los resultados obtenidos son sensiblemente mejores, si bien la variable más relevante sigue siendo la renta.

Sorprendentemente hay un producto para el que los resultados son muy diferentes. Así, el poseer o no microondas viene explicado por el poder de compra de la unidad familiar y por el nivel de formación de la mujer. Este hecho es fácilmente explicable si consideramos que se trata de un producto de consumo duradero con una gran variedad de precios donde, por tanto, el nivel de ingresos de la familia no es un factor determinante en su adquisición.

La compra del lavavajillas en el caso de mujeres que trabajan fuera del hogar se explica por el nivel de ingresos de la familia, la formación de ambos cónyuges, el poder de compra de la unidad familiar y el nivel de ocupación de la mujer.

Análisis logit

El análisis logit es una extensión del análisis de regresión cuando la variable dependiente toma los valores 0 ó 1, en función de que pertenezca a una clase dentro de la población. La estimación de la variable dependiente en función de las independientes permite predecir cuál es la probabilidad de que un elemento de la población con unas características determinadas manifieste un cierto comportamiento; por ejemplo, en este caso, la probabilidad de que adquiera el producto.

Para estimar el modelo logit se utiliza el programa TSP42B.

(2) Categoría 1: empresario-cargo directivo y profesión liberal. Categoría 2: cargo de responsabilidad intermedia y resto de trabajadores y autónomos. Categoría 3: ama de casa, en paro y pensionista.

(3) Categoría 1: sin estudios. Categoría 2: Estudios primarios y Bachiller. Categoría 3: titulación universitaria.

CUADRO 5.L.3. Resultados del análisis discriminante para diferenciar entre familias compradoras y no compradoras (mujer trabaja fuera del hogar)

Variable	*REN*	*NAU*	*TUF*	*TH*	*TM*	*EH*	*EM*	*Número compradores*	*Porcentaje clasificados correctamente*
Frigorífico**	—	—	—	—	—	—	—	—	—
Microondas	1.º 21,9	—	—	—	—	—	—	106	50
Lavavajillas	1.º 19,1	—	—	—	2.º 9,6	—	—	88	73
Vitrocerámica	1.º 29,4	—	—	—	—	—	—	65	63
Automóvil	1.º 9,92	*	—	—	—	—	—	137	74

* Variable no incluida.
** No discrimina ninguna variable.

CUADRO 5.L.4. Resultados del análisis discriminante para diferenciar entre familias compradoras y no compradoras (mujer no trabaja fuera del hogar)

Variable	*REN*	*NAU*	*TUF*	*TH*	*TM*	*EH*	*EM*	*Número compradores*	*Porcentaje clasificados correctamente*
Frigorífico**	1.º 10,3	—	—	—	—	—	—	180	46
Microondas		1.º 13,5	—	—	—	—	2.º 6,27	36	65
Lavavajillas	1.º 58,7	3.º 4,1	—	—	4.º 13,7	2.º 13,7	3.º 4,37	81	78
Vitrocerámica	1.º 20,8	—	—	—	—	—	—	69	68
Automóvil	—	*	—	1.º 27,9		—	—	164	79

* Variable no incluida.
** No discrimina ninguna variable.

Los resultados del análisis efectuado para toda la muestra se recogen en el Cuadro 5.L.5. El producto elegido ha sido el lavavajillas, ya que el resto únicamente eran explicados por la renta.

Para interpretar los resultados se toma en consideración, fundamentalmente, el signo de los parámetros de las variables estadísticamente significativas.

Observando el cuadro, puede verse como a mayor nivel de renta hay una probabilidad más alta de adopción. El resto de variables con coeficientes estadísticamente significativos son el grado de formación y el status de trabajo de la mujer, variables que han sido previamente categorizadas.

Nuestro objetivo en el futuro cercano es incorporar nuevos productos (secadora, robot doméstico...) que nos permitan detectar un mayor grado de discriminación.

CUADRO 5.L.5. Resultados del análisis Logit para el lavavajillas

Variable	*Coeficientes*	*t de Student*
REN	0,694-06	6,5650
EM categoría 1	–4,2498	4,0654
EM categoría 2	–2,5380	–6,0772
EM categoría3	–2,0713	–3,8812
TM categoría 1	–1,94586	–3,0939
μTM categoría 2	0,1941	0,5488

$R^2 = 0{,}34$; LR = 104,28.

5. BIBLIOGRAFÍA

BRYANT, W. K. (1988): «Durables and Wives Employment Yet Again», *Journal of Consumer Research,* 15 june), 37-47.

BURNS, D. J. (1992): «Husband-Wife Innovativa Consumer Decision Making: Exploring the Effect of Power Family», *Phycology & Marketing,* vol. 9 (33) (may/june), 175-189.

CHANKON, K. (1989): «Working-Wifes time-saving tendecies: durable ownership convenience food», *Journal of Economics Psychology.* 10 (nov.), 391-409.

DOUGLAS, S. P. (1976): «Cross-National Comarisons and Consumer Steoreotypes: A Case Study of Working and Non-Working Wives in the U. S. and France». *Journal of Consumer Research,* 3 (june), 12-20.

LOVINGOOD, R. P., and MCCULLOUNGH, J. L. (1986): «Appliance Ownership and Household Work Time», *Home Economies Reseorch Journal,* 14, 3 (march), 326-335.

NICHOLS, S. Y., and FOX, K. D. (1983): «Buying Time and Saving Time: Strategies for Managing Household Production», *Journal of Consumer Research,* 10 (september), 197-208.

OROPESA, R. S. (1993): «Femila Labor Force Participation and Time Saving Household Technology: A Case study of the Micowave from 1978 to 1989», *Journal of Consumer Research,* 19 (march), 567-579.

REILLY, M. D. (1982): «Working Wife and Convenience Consumption», *Journal of Consumer Research,* 8 (march), 407-419.

SCHANINGER, Ch. M., and ALLEN, C. T. (1981): «Wife's Occupational Status as a Consumer Behavior Construct», *Journal of Consumer Research,* 8 (september), 189-196.

SOBERON-FERRER, R., and DARDIS, R. (1991): «Determiants of Houdehold Expenditures for Services» *Journal of Consumer Research,* 17 (march), 385-397.

STROBER, M. R., and WEINBERG, Ch. B. (1980): «Strategies Used by Working and Nonworking Wives to Reduce Time pressures», *Journal of Consumer Research,* 6 (march), 338-348.

WEINBERG, Ch. B., and WINER, R. S. (1983): «Working Wives and Mayor Family Expenditures: Replication and Extension», *Journal of Consumer Research,* 10 (september), 259-263.

LECTURA COMPLEMENTARIA

Las estrategias de captación y retención de clientes y la lealtad. Un análisis de segmentación

SANDRA CAVERO BRÚJULA
JAVIER CEBOLLADA CALVO*
Universidad Pública de Navarra

Revista Española de Investigación de Marketing-ESIC, n.º 1
Septiembre, 1997

En este trabajo se analizan algunas pautas para determinar una asignación óptima de recursos entre las estrategias de captación y retención de clientes. Se presenta una metodología que permite segmentar a los consumidores según su grado de lealtad y estimar la respuesta diferencial de los segmentos a la actividad de marketing de la empresa. Se realiza una aplicación empírica sobre el mercado de los detergentes y se extraen conclusiones interesantes acerca de los aspectos mencionados, así como de la medición del resultado intangible de la actividad comercial y del valor de marca. Otras conclusiones interesantes encontradas se refieren al valor de las marcas nacionales frente a otro tipo de marcas y la importancia creciente de los grandes detallistas en la gestión de la retención de clientes.

Palabras clave: *lealtad; segmentación; estrategias de captación y retención de clientes; marcas nacionales y marcas de distribuidor; análisis cluster; logit multinomial; análisis de correspondencias.*

1. INTRODUCCIÓN

La segmentación de los mercados ha sido siempre un tema de interés central para el marketing. Es sabido que cuando en el mercado existen consumidores con preferencias distintas sobre los productos o que cuando éstos responden de manera distinta a las acciones comerciales, las empresas pueden incrementar sus beneficios si llevan a cabo una estrategia comercial diferenciada para los distintos segmentos. En un artículo reciente, Blattberg y Deighton (1996) exponen una nueva manera de segmentar cuando manifiestan que los objetivos del marketing pueden englobarse en dos: captar nuevos consumidores y retener los actuales[1]. Según ambos autores, debería considerarse incluso desarrollar planes de marketing diferentes para cada tipo de clientes, y cada empresa debería determinar, a nivel de cada una de sus marcas, el balance óptimo entre los esfuerzos destinados a cada objetivo.

Hablar de retención de dientes es hablar de lealtad, entendida en sentido amplio: un cliente leal es aquel que permanece con la empresa durante un largo tiempo, o dicho en otras palabras, aquel que en sucesivas compras elige de manera repetida los productos de la empresa. A la hora de determinar la asignación óptima de recursos entre las estrategias de captación y retención de clientes, ha de considerarse la rentabilidad comparada de ambas. Aaker (1994) recuerda que, en general, la inversión necesaria para vender a los clientes actuales es menor que la necesaria para los clientes nuevos. Otros autores han mostrado asimismo que la inversión en promociones para conseguir una venta es menor entre los clientes leales, y que éstos generan un mayor ingreso que los no leales debido a la menor elasticidad - precio de su demanda (Krishnamurthi *et al.*, 1995). Por tanto, conocer la rentabilidad de los clientes leales y no leales es

* Departamento de Gestión de Empresas, campos de Arrosadía, 31006 Pamplona. Internet: cebollada@upna.es tel: 948-169736, fax: 948-169404. Los autores agradecen a V. Salas la ayuda prestada y a J. M. Múgica los consejos recibidos. También agradecen a Dympanel, S. A., la cesión de los datos utilizados.

[1] En inglés *acquiring y retaining*, respectivamente.

una información muy importante para determinar la asignación óptima de recursos entre las estrategias de captación y retención de clientes.

Por otro lado, la lealtad es considerada como una buena medida del valor de marca (Blackstone 1990, Aaker 1994) y del resultado de la empresa. Como señalan Fornell y Wernerfelt (1987), la última consecuencia de la satisfacción del consumidor es la lealtad, debido a su carácter de variable *proxy* de la rentabilidad de la empresa (Reichheld y Sasser 1990). En la misma línea y analizando la productividad del área de marketing de las empresas, Sheth y Sisodia (1995) argumentan que una de las causas por las que en muchas empresas se duda de su rentabilidad es que los resultados del marketing tienen un alto grado de intangibilidad y por tanto son difícilmente observables y medibles. Según estos autores, los resultados deberían ser medidos no sólo en términos cuantitativos sino también cualitativos, incluyendo el nivel de satisfacción y lealtad de los clientes y el valor de las marcas de la empresa. Sheth y Sisodia recomiendan medir el *output* del marketing en términos adquisición y retención de consumidores. El argumento que sustenta tal proposición es que la tasa de retención o grado de lealtad informa sobre qué parte de los beneficios corrientes del año (los que proporcionan los estados financieros clásicos como la cuenta de resultados y el balance de situación) serán estables a lo largo del tiempo. Por tanto, como hemos ido viendo, la medición de la lealtad constituye un elemento central en la medición del resultado de la actividad de marketing en general, y en particular para determinar la asignación óptima de recursos entre los objetivos de adquisición y retención de clientes.

En cuanto a la medición de la lealtad aparecen en la literatura dos aproximaciones (Jacoby y Chesnut 1978): la primera pone énfasis en las actitudes y la segunda en el comportamiento. Según la primera, la lealtad toma la forma de un compromiso psicológico del consumidor con la marca, que se manifiesta en una actitud positiva y en una intención de recompra. Según la segunda, la lealtad se manifiesta como la recompra persistente de una marca a lo largo del tiempo. La medición se realiza por tanto a partir de la secuencia histórica de compras del consumidor. Los favorables a esta aproximación argumentan en contra de la primera que, con frecuencia, y por diversos motivos, la intención no se transforma en compra efectiva y que medir el comportamiento es más objetivo que medir las actitudes. Los partidarios de la primera aproximación manifiestan en contra de la segunda que no contempla información importante acerca de las causas de la lealtad. Como intento de solución de compromiso, una tercera aproximación intenta recoger el verdadero y complejo carácter mixto de la lealtad, sin que haya tenido éxito debido a la dificultad de hacer operativa correctamente tal aproximación y a la de encontrar datos de ambos aspectos conjuntamente (véase por ejemplo Olson y Jacoby 1971).

En los últimos años, en paralelo a la creciente disponibilidad de datos del panel de consumidores, la medición a partir de la secuencia de compras ha experimentado un gran auge en la literatura, siendo el artículo de Guadagni y Little (1983) uno de los pioneros. Este tipo de datos permite, además de la tasa de retención, medir el efecto de la lealtad en el beneficio mediante la disminución de la elasticidad-precio a que da lugar. Pero de manera cuando menos sorprendente, los artículos de la literatura que analizan este aspecto obtienen resultados contradictorios. Como muestra, Massy y Frank (1965) no encontraron diferencias significativas en las elasticidades-precio entre familias con diferentes grados de lealtad. Frank (1967) tampoco encontró evidencia empírica para validar la hipótesis de que los consumidores leales se diferenciaban de los no leales en su respuesta a actividades promocionales. Starr y Rubinson (1978) hallaron, en cambio, una relación inversa entre lealtad y elasticidad-precio, comparando la magnitud de ésta última entre marcas que tenían diferentes grados de lealtad. Por otro lado Neslin *et al.* (1985) analizaron el efecto de las promociones en el aumento de las ventas y la diferencia entre grupos de consumidores leales y no leales y encontraron que el efecto dependía del producto estudiado: mientras que en el papel higiénico no hallaron diferencias, en el café los leales incrementaban la cantidad comprada más que los no leales. McCann (1974) concluyó, en cambio, que cuando los consumidores centran sus compras en unas pocas marcas son más sensibles a los cambios en el precio que cuando las reparten entre varias.

Krishnamurthi y Raj (1991) consideran que los aparentemente inconsistentes resultados de la literatura se deben a que se han analizado diferentes etapas en la decisión de compra: mientras que Starr et al. (1978) se basan en la decisión de elección de marca, Neslin *et al.* (1985) lo hacen en la decisión de cuantas unidades. Argumentan también que sólo usando datos desagregados a nivel de observaciones de compras pueden distinguirse estas dos etapas; en caso de utilizar datos agregados a nivel de cuota de mercado o ventas, las conclusiones dependerán de cuál de las dos decisiones domina. Estos autores analizan las dos decisiones por separado y encuentran una relación inversa entre elasticidad - precio y lealtad en la decisión de elección de marca pero una relación positiva en la decisión de cantidad.

Sin embargo, este trabajo es criticable por dos motivos. En primer lugar, desestima para su estudio a los consumidores que no cumplen una serie de requisitos de intensidad de compra y dedicación a las marcas líderes, lo que con seguridad convierte a la muestra de individuos en sesgada por selección defectuosa. En segundo lugar, sus resultados están basados en la comparación entre compras (y no individuos) leales y no leales. Según esta clasificación un mismo consumidor puede realizar a lo largo del tiempo unas compras calificables como leales y otras calificables como no leales. Esta clasificación se establece sobre el vector de probabilidades de elección de cada individuo, que está formado por la proporción de veces que el consumidor ha comprado cada una de las marcas hasta su última compra y que se actualiza cada vez que se realiza una nueva compra. Desde nuestro punto de vista, esta manera de clasificar es difícil de justificar conceptualmente, ya que sitúa la lealtad en las compras y no en el consumidor. Además no permite segmentar el mercado por grupos de compradores leales y no leales, y por tanto no es útil para las empresas en la gestión de sus marcas.

Por tanto, en nuestra opinión, hasta el momento no se ha analizado con el suficiente rigor empírico y conceptual la relación entre lealtad y elasticidad-precio. Cuando se han utilizado datos agregados, se han enmascarado dos decisiones que pueden tener efectos contrarios, sin conocerse cuál es el que domina. Por otro lado, los trabajos con datos individuales carecen del rigor necesario. A lo ya comentado sobre el trabajo de Krishnamurthi y Raj (1991), puede añadirse que el análisis de Starr y Rubinson (1978), al establecer comparaciones entre marcas, no considera que éstas pueden tener otros factores distintivos además del índice de lealtad, por lo que las diferencias halladas no pueden asignarse con seguridad a la lealtad. El estudio de Neslin *et al.* (1985), se centra en la decisión de cantidad pero no considera la decisión de elección de marca que es generalmente más importante.

En este trabajo pretendemos armonizar los aspectos mencionados a lo largo de esta sección. Presentamos un procedimiento que, en primer lugar segmenta el mercado en grupos de consumidores leales y no leales, superando así el problema que presenta el trabajo de Krishnamurthi et al. (1991) que clasificaba las compras y no los consumidores. En segundo lugar, se identifican las características sociodemográficas y de hábitos de compra de los consumidores, de manera que esa segmentación pueda ser efectivamente llevada a cabo por las empresas. Además, estimamos la elasticidad - precio de ambos grupos de consumidores para contrastar si, además de mayor tasa de retención, los consumidores leales pueden soportar un mayor incremento de precio que los no leales. De esta manera, el procedimiento no solo proporciona una tipología de los consumidores sino que permite que la segmentación pueda ser implementada como estrategia, recogiendo la distinción de Sarabia y Munuera (1994). Ambos aspectos, la identificación de los segmentos y la estimación de su diferente respuesta al marketing, permite a las empresas evaluar planes de marketing alternativos para la captación y retención de clientes. El estudio se realiza analizando datos de un panel de hogares[2] con información del mercado de los detergentes.

En el siguiente apartado se lleva a cabo una aplicación empírica mencionada. En primer lugar describimos el mercado. A continuación realizamos una segmentación que nos permitirá extraer información útil para la empresa a la hora de articular las estrategias de captación y retención de clientes. Más concretamente, analizamos la lealtad a tipos o variedades de marcas de detergentes, en vez de a marcas comerciales a nivel individual. Estas variedades están formadas por marcas relativamente homogéneas que comparten entre si atributos que las hacen susceptibles de agregación, como formato[3] del paquete, precio, respaldo publicitario, y estrategia comercial. En el Apartado 4 se exponen las conclusiones más importantes del trabajo y los comentarios finales.

[1] Baldinger (1992) propone utilizar datos de panel para determinar los segmentos del mercado según su comportamiento hacia la marca a lo largo de su historial de compras.

[2] Por formato entendemos tamaño o peso.

2. METODOLOGÍA Y APLICACIÓN EMPÍRICA

En este apartado se describe el mercado, que posteriormente se analiza. En primer lugar, se lleva a cabo una segmentación de los consumidores según el grado de lealtad y posteriormente se caracteriza a los segmentos encontrados. Finalmente se analiza su respuesta diferencial al marketing.

2.1. Los datos y el mercado

Para analizar el mercado mencionado utilizamos una base de datos proveniente de un panel de hogares españoles comercializado por la empresa Dympanel, S. A. Este panel es representativo a nivel áreas Nielsen, está compuesto por alrededor de 4.000 hogares (universo: 11 millones de hogares en España) y se renueva un 25 por ciento cada año. Se han seleccionado las 4.156 compras de detergente para ropa, en polvo, normal, para lavadora automática realizadas por los 235 hogares miembros del panel residentes en el área Nielsen de Barcelona ciudad y su área metropolitana (en la que el panel es representativo) durante los años 1992 y 1993. Esta doble selección producto-zona geográfica configura un mercado bien delimitado Para contar con un producto más homogéneo se han seleccionado las compras de paquetes de 2 y de 4 kg de peso, quedando de esta manera 3.753 observaciones para su análisis. Cada observación consta del número de identificación del hogar sus principales características sociodemograficas (categorizadas por tramos) y diversa información acerca del producto comprado y del acto de compra.

El conjunto de elección al que se enfrentan los hogares está formado por 77 alternativas que se diferencian fundamentalmente en su marca comercial y en el formato del paquete. Por razones operativas (es materialmente imposible trabajar con todas las alternativas) y de interés en el análisis de la lealtad hacia variedades de marcas, agruparemos las alternativas en 8 variedades, en función del tipo de marca (segundas marcas, marcas nacionales y marcas de distribuidor[4]; dentro de las de distribuidor distinguimos entre las de hipermercado y las de supermercado para homogeneizar más internamente las variedades) y del formato (2 y 4 kg). Las marcas nacionales son líderes en el mercado (las 10 primeras marcas en cuota de mercado son nacionales), tienen un nombre comercial conocido, realizan enormes inversiones publicitarias (que las lleva a situarse entre las primeras en los rankings de anunciantes) y un precio sensiblemente superior al resto (véase Cuadro 5.L.1). Las marcas de distribuidor son propiedad y llevan el nombre del establecimiento comercial en el que se venden, no realizan campañas publicitarias (salvo las que realizan genéricamente los establecimientos) y su cuota y precio están muy por debajo de las nacionales. Las marcas segundas tienen nombres comerciales no conocidos y al igual que las de distribuidor no se anuncian y su precio y cuota son similares a estas últimas.

[4] Esta clasificación de las marcas puede encontrarse en cualquier libro de texto de marketing.

CUADRO 5.L.1. Cuota de mercado y precios de las variedades de elección

Formato	*Tipo*	*Cuota (%)*	*Precio medio (Ptas/paquete)*
2 kg	Segundas marcas	2,3	636,0
"	Nacionales	36,9	747,8
"	Distrihuidor-hiper	0,7	508,1
"	Distnbuidor-super	4,3	527,2
4 kg	Segundas marcas	5,8	528,7
"	Nacionales	42,4	722,9
"	Distribuidor-hiper	1,9	570,8
"	Distribuidor-super	5,7	582,3
Total		100,0	701,3
	2kg	44,22	718,2
	4kg	55,8	686,6
Total		100,0	701,3
	Segundas marcas	8,1	564,3
	Nacionales	79,3	737,8
	Distribuidor-hiper	2,6	556,0
	Distribuidor-super	10,0	558,7
Total		100,0	701,3
Número medio de compras por hogar en el período			16,4

Los detergentes en formatos de 2 kg son concentrados y tanto su capacidad de lavado como su precio son similares a los de los paquetes de 4 kg. Por otro lado, análisis llevados a cabo por diversas asociaciones de consumidores han llegado a la conclusión de que la calidad de las diferentes marcas de detergentes, sean nacionales, de distribuidor o segundas es, en general, similar[5]. En el Cuadro 5.L.1 se muestra una descripción de los precios y cuotas de las 8 variedades.

2.2. Segmentación de los consumidores a través de la lealtad

Según Grover y Srinivasan (1987) «un segmento es un grupo de consumidores definido en términos de probabilidad de elegir las diferentes marcas de un mercado. Un segmento está representado por un vector homogéneo de probabilidades de elección y la diferencia entre dos segmentos estriba en que sus correspondientes vectores son diferentes». Nuestro interés en este apartado consiste en encontrar segmentos de hogares leales y no leales mediante la identificación de grupos homogéneos de hogares con probabilidades de compra parecidas entre sí y distintas a las de otros grupos. Una buena estimación del vector de probabilidades de elección de un consumidor sobre las 8 variedades descritas anteriormente viene dada por la proporción de veces que el hogar ha comprado cada una de ellas sobre el total de compras que ha realizado. Más formalmente, si N_{ij} es el número de veces que en la base de datos el consumidor i ha comprado la variedad j y N_i es el número total de compras realizadas por i, $h_{ij} = N_{ij}/N_i$ es la probabilidad estimada de que i elija j en una compra cualquiera. Si $h_i = (h_{i1}, h_{i2}, ..., h_{i8})$, es el vector o perfil de probabilidades de compra del consumidor i sobre las variedades $j = 1, 2, ..., 8$, un hogar puede ser considerado leal a la marca k si h_{ik} es superior a cualquier otro h_{ij}. Nosotros consideraremos que un individuo i es leal a la variedad j si $h_{ij} > 0,5$, asegurándonos de esta forma que es leal a una sola marca[6]. Para agrupar a hogares con perfiles de probabilidad parecidos hemos empleado la técnica de análisis cluster sobre los vectores h_i.

Los resultados de la segmentación aparecen en el Cuadro 5.L.2. Se han identificado 8 segmentos de hogares, para cada uno de los cuales se observa el perfil de probabilidad medio y su tamaño (en número de hogares y en ventas). De los 8 segmentos encontrados, 7 están formados por hogares considerados leales a alguna de las variedades (la variedad marcas de hipermercado de 2 kg no tiene ningún hogar con probabilidad de compra mayor que 0,5) y uno por hogares no leales a ninguna. Se observa que las marcas nacionales superan a las demás tanto en cuota de mercado como en grado de lealtad (medida ésta como la probabilidad de elección

[5] Análisis aparecidos en las revistas OCU-Compra Maestra (1995), Ciudadano (1992 y 1994) y Eroski (1994).

[6] Es interesante resaltar la objetividad de esta medida: todo hogar es inequívocamente clasificado como leal o no leal. En la literatura aparecen índices más o menos exigentes en cuanto al número de repeticiones o porcentajes de compras necesario para considerar a un individuo como leal, pero 0,5 es el punto más utilizado.

CUADRO 5.L.2. Segmentos de hogares leales a cada marca y no leales. Perfil medio de probabilidades de elección, número de hogares y número de compras. Información para cada segmento y para el conjunto de todos los hogares. Entre paréntesis, y debajo de la marca objeto de lealtad, probabilidades de compra mínima y maxima de esa marca entre los hogares qué componen el segmento

Leales a	*Variedades de marcas*										
	4kg				*2kg*				*Suma*	*%*	*%*
	Seg.	*Nac.*	*Hiper*	*Super*	*Seg.*	*Nac.*	*Hiper*	*Super*		*Hogares*	*Compras*
Segundas 4kg	0,77 (0,52-1,00)	0,18	0,00	0,00	0,01	0,03	0,00	0,00	1,00	5,0	6,5
Nacionales. 4 kg	0,03	0,84 (0,0,58-1,00)	0,02	0,02	0,00	0,09	0,00	0,00	1,00	33,0	39,2
Disttrib-Hiper 4 kg	0,00	0,11	0,76 (0,63-0,91)	0,08	0,00	0,04	0,00	0,00	1,00	1,3	0,7
Disttrib-Super 4 kg	0,07	0,11	0,00	0,78 (0,58-1,00)	0,00	0,01	0,00	0,01	1,00	1,8	4,1
Segundas 2kg	0,00	0,03	0,00	0,02	0,94[17]	0,00	0,00	0,02	1,00	0,0	1,8
Nacionales 2 kg	0,00	0,08	0,00	0,00	0,00	0,88 (0,58-1,00)	0,00	0,02	1,00	34,0	31,7
Disttrib-Super 2 kg	0,02	0,00	0,00	0,11	0,01	0,03	0,02	0,81 (0,51-1,00)	1,00	2,9	2,1
No leales	0,06	0,30	0,06	0,11	0,02	0,31	0,04	0,10	1,00	19,0	13,9
Perfil medio de todos los hogares	0,06	0,42	0,02	0,06	0,02	0,37	0,01	0,04	1,00	100,0	100,0

[17] Sólo hay un hogar en este segmento.

de la marca a la que el segmento es leal[7]. El segmento leal a las nacionales concentradas supera al de las no concentradas en grado de lealtad y en número de clientes, pero no así en ventas, lo que significa que el segmento de leales a las nacionales de 4 kg está formado por hogares con una mayor tasa de consumo. El tercer segmento en importancia en cuota de mercado es el de los leales a las segundas marcas de 4 kg, que por otro lado tiene un grado de lealtad substancialmente menor. A continuación está el segmento de los leales a las marcas de supermercado de 4 kg, que además esta forma- do por hogares con una altísima tasa de consumo (1,8por ciento de los hogares, 4,1 por ciento de las ventas). Los restantes segmentos tienen cuotas de mercado substancialmente menores. Se observa en general que los consumidores leales a las variedades de 4 kg tienen una mayor tasa de consumo que los leales a las de 2 kg y ambos mayor que los no leales. En la Figura 5.L.1 se ha dibujado, a título representativo, los vectores de probabilidades de elección medios de los leales a las marcas nacionales de 2 y 4 kg, del segmento no leal y el medio de todo el mercado. Se observa con claridad como los segmentos leales tienen un perfil que se hace máximo en la variedad a la que son leales, mientras que el segmento no leal reparte más sus compras y tiene un perfil mucho más plano.

A partir de este análisis de segmentación también se puede conocer el patrón de sustitución entre variedades (Figura 5.L.2). Todas las variedades, excepto las marcas de supermercado de 2 kg, tienen las nacionales de 4 kg como segunda opción. En cambio, las nacionales de 4 kg tienen las nacionales de 2 kg como segunda alternativa y las de supermercado de 2 kg tienen el mismo tipo de marca en 4 kg. Una posible interpretación a este hecho es que los compradores de las marcas no nacionales compran las nacionales de 4 kg cuando están en oferta.

Otro aspecto interesante a señalar es que el segmento de los no leales está compuesto por consumidores que, a lo largo del periodo de dos años, cambiaron más que los demás segmentos del formato de 4 al de 2 kg en las marcas nacionales, lo que parece indicar que los consumidores no leales están más predispuestos a las innovaciones (en este caso de una reducción en el peso del paquete paralela a la sustitución del detergente clásico por uno concentrado). Esto se observa en el Cuadro 5.L.3, donde aparecen las probabilidades de elección medias de ambas variedades por semestres.

2.3. Caracterización de los hogares pertenecientes a los segmentos

Para las empresas no sólo es interesante saber que los consumidores son heterogéneos en sus probabilidades de com-

[7] Aunque el perfil medio de los leales a las segundas marcas de 2 kg muestra una probabilidad de comprar esa marca de 0,94, está formado por un solo consumidor por lo que no es representativo.

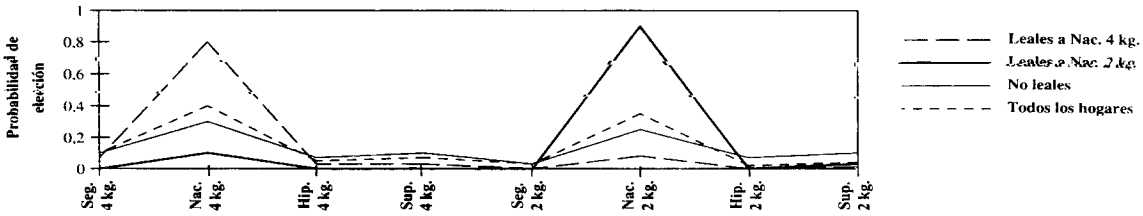
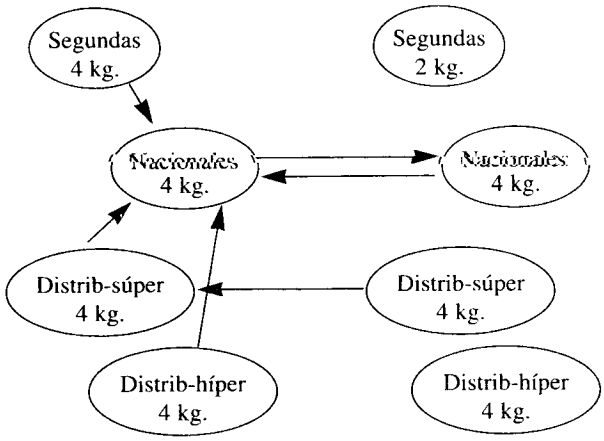


FIGURA 5.L.1. Perfil de elección de los segmentos.



Nota: las marcas en el origen de la flecha tienen como segunda opción las marcas en el destino de la flecha.

FIGURA 5.L.2. Patrones de sustitución entre marcas.

CUADRO 5.L.3. Probabilidades de elección medias del segmento de no leales hacia las marcas nacionales de 2 y 4 kg. Evolución por semestres

	SEM.1	*SEM.2*	*SEM.3*	*SEM.4*	*Media 2 años*
Nacionales 4 kg	0,43	0,33	0,24	0,20	0,30
Nacionales 2 kg	0,20	0,40	0,38	0,47	0,31

pra y, como consecuencia, en su grado de lealtad y que pueden agruparse en segmentos. Para poder aprovechar esta información llevando a cabo estrategias de marketing diferenciadas en un mercado de consumidores como el de este producto, necesitan saber cuál es perfil de cada segmento, en términos de características sociodemográticas, de hábitos de compra y/o de uso del producto.

En la literatura se han analizado distintas variables relacionadas con la lealtad. Cebollada (1996) menciona que los hogares con más miembros son mas leales que los que tienen menos y apuesta que una posible razón es su mayor tasa de consumo, y que por tanto la verdadera fuente de la lealtad es el aprendizaje debido a la experiencia con el producto. En el mismo trabajo, se observa que entre las compras realizadas en super e hipermercados el grado de lealtad es menor, relacionándose este hecho con la mayor variabilidad de precios y actividad promocional en este tipo de establecimientos, así como con su régimen de autoservicio que facilita a los consumidores la comparación entre marcas sin influencia del dependiente. La base de datos a nuestra disposición nos proporciona, además del número de miembros del hogar, tasa de consumo y tipo de tienda donde se ha realizado la compra, información acerca de la edad del ama de casa y la clase social del hogar. Con respecto al lugar donde se realiza la compra, hemos llevado a cabo un análisis cluster de la elección de tipo de establecimiento comercial similar al realizado para la elección de la variedad de marca y hemos identificado dos segmentos de hogares, en función del tipo de establecimiento de compra preferido por éstos[8]. El resultado de este análisis, que aparece en el Cua-

[8] El tipo de tienda principalmente elegido por el consumidor ya ha sido utilizado otras veces como variable de segmentación, como por ejemplo en Hernández, Munuera y Ruiz (1995). Dado que en este trabajo nuestro interés no se centra en la elección de establecimiento comercial no daremos más detalles acerca de este análisis. Si el lector está interesado en el mismo, puede contactar con los autores.

CUADRO 5.L.4. Segmentación de los hogares según el tipo de tienda por el que muestran preferencia. Perfil de probabilidad de elección de tipo de establecimiento

Segemento	*Hyper y supermercados*	*Resto de tiendas*	*Suma por filas*	*N hogares*	*% hogares*
1	0,94	0,06	1,00	194	82,6
2	0,25	0,75	1,00	41	17,4
Media de todos los hogares	78,6	21,4	1,00	235	100,0

dro 5.L.4, proporciona dos segmentos claramente diferenciados: los hogares que compran principalmente en hiper y supermercados, y los que compran principalmente en otro tipo de tiendas.

Para identificar las características de los consumidores leales y no leales se ha realizado un análisis de correspondencias múltiple (Greenacre 1993) de las variables a nuestra disposición que presumiblemente pudieran estar relacionadas con la lealtad. La representación gráfica de este análisis aparece en la Figura 5.L.3, y se observa que la categoría «no leal» está más correlacionada con el tamaño familiar pequeño, preferencia por comprar en super e hipermercados, menor tasa de consumo («light user» en la figura), menor edad y clase social más alta. Lo contrario puede decirse de la categoría «leal»[9]. Para confirmar este análisis exploratorio, se ha realizado un análisis logit binomial de la probabilidad de que un hogar sea considerado como leal en función de las variables anteriormente mencionadas. En el Cuadro 5.L.5 se muestra un análisis logit binomial que identifica las características y la forma en que afectan a la probabilidad de que el hogar sea leal frente a que no lo sea, así como la definición de las variables. Según este modelo la probabilidad de que un hogar i sea leal puede expresarse como función de la utilidad V_i del individuo de la siguiente manera

$$Pi(LEAL) = \frac{\exp[V_i(LEAL)]}{\exp[V_i(LEAL)] + \exp[V_i(NO\ LEAL)]} \quad (1)$$

donde

$$V_i(LEAL) = \beta_0 + \beta_1 EDAD_i + \beta_2 TAMAÑO_FAMILIAR + \\ + \beta_3 CLASE_SOCIAL_i + \\ + \beta_4 GRAN_CONSUMIDOR_i + \\ + \beta_1 HIPER_SUPER. \quad (2)$$

Las variables representan la edad del ama de casa, el número de miembros del hogar, la clase social, si consume el producto por encima de la tasa media de todos los hogares (8,2 compras/año) y si es comprador habitual de hiper y supermercados (según la segmentación realizada con anterioridad). Estas variables están definidas con más exactitud en el propio Cuadro 5.L.5. A continuación comentamos los resultados obtenidos.

La variable edad del ama de casa no aparece como significativa ($t = 0,22$) y por tanto, no podemos concluir que la edad sea una variable relacionada con la lealtad, al menos en este producto. El efecto del número de miembros del hogar es significativo y positivo (coef. = 0,38, $t = 2,06$), lo que indica que las familias más grandes son más leales. La clase social no es significativa en esta especificación, pero en una en la que es la única explicativa, pasa a serlo (coef. – 0,29, $t = -1,99$) y el signo negativo indica que ser de clase social baja aumenta la probabilidad de que el hogar sea leal. La variable que refleja la tasa de consumo es la más significativa y su efecto es el más positivo (coef. = 1,13, $t = 2,86$), indicando que los hogares que compran más veces son más leales, y que este efecto es el más importante de entre todos los analizados. Si asumimos que el efecto del aprendizaje es mayor en estos hogares, se refuerza la hipótesis de que una de las causas de la lealtad es la experiencia acumulada con el producto. Por último, el efecto de la variable de preferencia por comprar en hiper y supermercados, es también significativo y, tal como hemos definido la variable, su signo negativo indica (coef. 5 20,69, t 5 21,77) que las familias que habitualmente compran en estos establecimientos son menos leales que las que lo hacen en otro tipo de tiendas. Esta relación puede explicarse por las características de este tipo de establecimientos: ofrecen un mayor número de marcas, realizan más actividades promocionales y al funcionar en régimen de autoservicio facilitan que el consumidor pueda comparar entre marcas sin influencia de dependiente alguno. Con todas las variables hemos realizado análisis individuales para comprobar que no hay problemas de multicolinealidad y los resultados no cambian en relación a los ya apuntados. Por tanto, a la vista del análisis anterior puede decirse que las familias grandes, de clase social media baja y baja, con una tasa de consumo alta y que compran poco en hiper y supermercados son más leales. Lo contrario puede decirse de los no leales.

Es también interesante caracterizar distintivamente a los hogares leales a cada una de las variedades, al igual que hemos hecho con los leales (en general, sin fijarnos a qué

[9] No nos extenderemos más en los comentarios de este análisis por restricciones de espacio. Si el lector está interesado en más detalles sobre este análisis puede contactar con los autores.

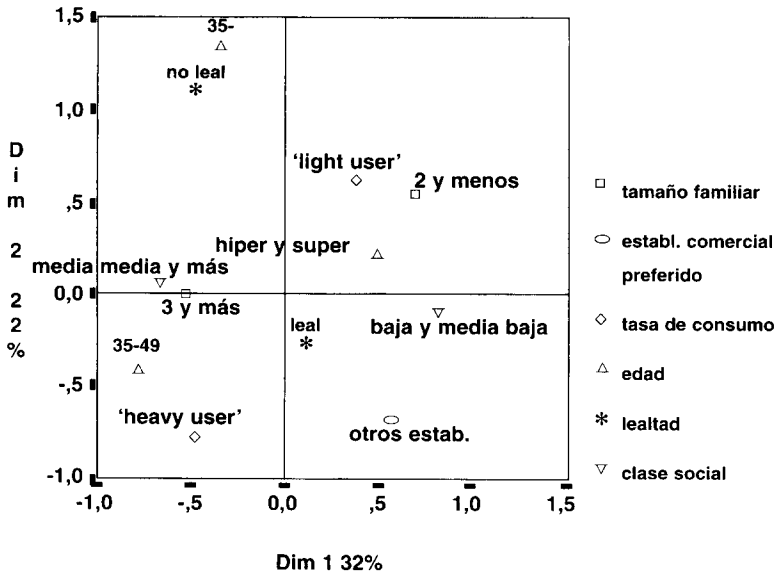


FIGURA 5.L.3. Análisis de correspondencias múltiples de las variables lealtad (leal/no leal), tamaño familiar, establecimiento comercial preferido, tasa de consumo, edad y clase social.

CUADRO 5.L.5. Análisis logit de los factores que afectan la probabilidad de que un hogar pertenezca al segmento de los leales

Variable dependiente: leal		
Variables independientes	*Coeficiente*	*t*
Constante	1,78	1,85
Edad	0,04	0,22
Tamaño familiar	0,38	2,06
Clase social	–0,26	–0,59
Gran consumidor	1,13	2,86
Hiper super	–0,69	–1,77
L(β)=-107,2	ρ^2_{adj}=0,07	
L(C)=-114,7	N=235	

Variables:

Leal	1 si el hogar pertenece a un segmento leal; 0 en caso contrario.
Edad	Edad del ama de casa. 1: menos de 35, 2: de 35 a 49, 3: 50 o más.
Tamaño familiar	Número de miembros del hogar. 1: 4 o más, 0: de 1 a 3.
Clase social	Clase social del hogar. 1: alta, media alta y media, 0: media baja y baja.
Gran consumidor	1 si el hogar ha comprado más veces que la media; 0 en otro caso. Media = 8,2 compras/año
Hiper super	1 si el hogar pertenece al segmento de preferencia hacia hiper y supermercados; 0 en otro caso.

Notas: 1. La medida de bondad del ajuste (similar al R_2 en la regresión) es ρ^2_{adj} = 1–[(L(β)–k)/L(C)], donde L(β) es el valor de la función de verosimilitud del modelo en cuestión, L(C) la del modelo que solo especifica constantes y k la diferencia del número de coeficientes entre ambos modelos.

2. Las características sociodemográficas de los hogares vienen definidas en el panel mediante variables categóricas. En algunas variables las categorías se han reagrupado buscando la mayor capacidad discriminante y explicativa.

marca) frente a los no leales. Para ello se ha realizado otro análisis de correspondencias múltiple en el que la variable lealtad tiene 8 categorías (una por cada segmento de lealtad a las distintas variedades de marcas de detergentes más una

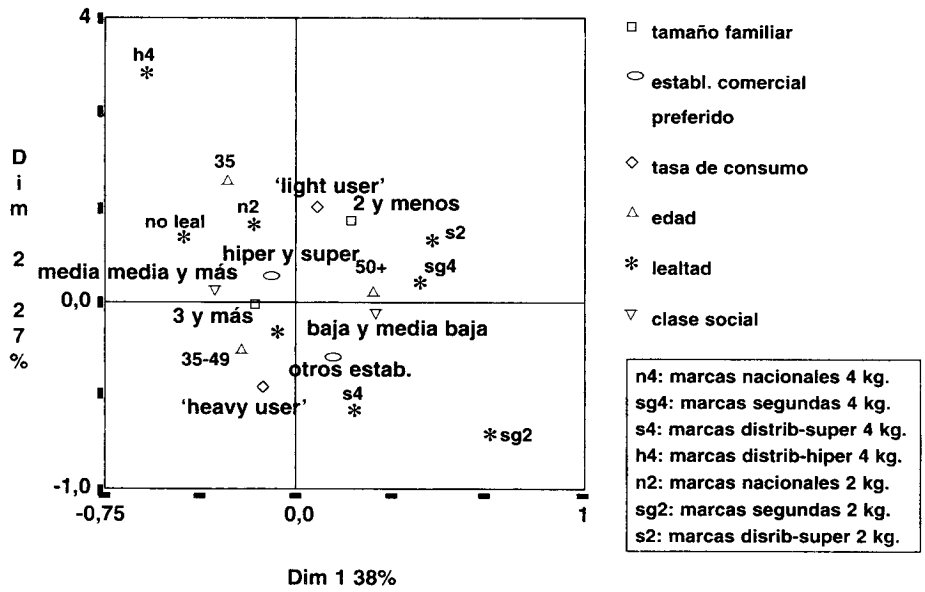


FIGURA 5.L.4. Análisis de correspondencia múltiple de las variables lealtad (a cada variedad de marcas), tamaño familiar, establecimiento comercial preferido, tasa de consumo, edad y clase social.

de no leales, conforme a los resultados del Cuadro 5.L.2) y cuya representación aparece en la Figura 5.L.4. Aunque los resultados de este análisis muestran ciertas asociaciones[10], éstas no han podido ser contrastadas significativamente mediante un análisis logit multinomial confirmatorio como el realizado a través de las ecuaciones (1) y (2), por lo que no podemos confirmarlas. Una explicación puede ser que el tipo de producto que analizamos está poco segmentado, más allá de las diferencias observadas en el grado de lealtad.

2.4. El efecto de la lealtad en la elasticidad-precio

Una vez segmentado el mercado en consumidores leales y no leales y conocido su perfil sociodemográfico, de elección de establecimiento comercial para la compra y de intensidad de consumo, el siguiente paso es conocer hasta qué punto ambos segmentos responden de manera diferenciada a las variables de marketing. En concreto, vamos a analizar si ambos grupos de consumidores tienen diferente elasticidad - precio. Si esto es así, la empresa podría aumentar sus beneficios llevando a cabo una política de precios diferente para ambos grupos de consumidores.

Para estimar la elasticidad-precio mediremos el efecto del precio sobre la probabilidad de elección de las ocho variedades, diferenciando si la compra ha sido realizada por un leal o por un no leal. Planteamos un modelo de elección logit multinomial (Guadagni y Little 1983) en el que la utilidad (indirecta) que cada hogar percibe al elegir una de estas variedades depende del precio de la alternativa elegida y de la preferencia del hogar por esa variedad[11]. Para distinguir entre hogares leales y no leales incluimos un efecto interactivo multiplicando el precio por una variable dicotómica de lealtad, como explicamos a continuación.

La probabilidad de que un hogar i elija la variedad j en la ocasión t, $P(VARIEDAD_{it}(j) = 1)$, puede expresarse mediante un modelo logit multinomial como

$$P(VARIEDAD_{it}(j) = 1) = \frac{\exp [V_{it}(j)]}{\sum_{j=1}^{8} \exp [V_{it}(j)]} \qquad (3)$$

donde la variable dependiente es

$VARIEDAD_{it}(j) = \{1$ si la variedad comprada por el hogar i en la ocasión t ha sido la j; 0 en caso contrario$\}$, siendo $j = 1, ..., 8$ cada una de las variedades que forman el conjunto de elección.

[10] Según este análisis los leales a las marcas nacionales serían de clase media-alta y alta, más jóvenes y de menor tamaño familiar, siendo los leales al formato de 2 kg menos consumidores del producto y aduciendo más a los super y droguería y tiendas de alimentación que los leales al formato de 4 kg. Los leales a las segundas serían de más edad y de clase social más baja. Los leales a las marcas de hipermercado (sólo los hay a la de 4 kg) serían sobre todo más jóvenes y los leales a las de supermercado serían de familias más pequeñas, y de mayor edad.

[11] Esta variable de preferencia u otras semejantes son usadas habitualmente en la literatura desde el trabajo de Guadagni y Little (1983) con el objeto de dotar al modelo de un carácter dinámico. En este caso utilizamos una variable semejante a la que utilizan Krishnamurthi *et al.* (1991).

y las variables explicativas son

P_{jt} = precio de la variedad j, en pesetas por paquete, en la observación t [12],
$PREF_{ijt}$ = Porcentaje de veces que el hogar i ha comprado la variedad j hasta la ocasión $t-1$ (incluida),
$LEAL_i$ = {1 si el hogar i pertenece a un segmento leal; 0 en otro caso} y
P_{jt} 3 $LEAL_i$: efecto interactivo.

La función de utilidad toma la siguiente forma aditiva

$$V_{it(j)} = \beta_{0j} + \beta_1 P_{jt} + \beta_2 PREF_{ijt} + \beta_3 (P_{jt} LEAL_i) \quad (4)$$

donde b_{0j} es una constante específica de cada variedad que toma el valor 1 si se ha comprado la variedad j y 0 en otro caso. Por tanto, el coeficiente de precio para los consumidores no leales será b_1 y para los leales $b_1 + b_3$. Usaremos las 12 primeras semanas para inicializar la variable PREF por lo que nuestro periodo de estimación comprende 92 semanas y 3.305 observaciones. Para tener una referencia y realizar comparaciones estimaremos también una ecuación como la anterior pero sin efecto interactivo y que tomará la forma

$$V_{it(j)} = \beta_{0j} + \beta_1 P_{jt} + \beta_2 PREF_{ijt} \quad (5)$$

Los resultados de la estimación de las ecuaciones (4) y (5), así como los de una especificación con constantes únicamente[13] aparecen en el Cuadro 5.L.6. La significatividad estadística y él signo de los coeficientes b_1 y b_3 de la ecuación (4) nos muestra que la variable precio tiene un efecto negativo sobre la probabilidad de elección de los hogares no leales (–0,0024), siendo su efecto del mismo signo para los hogares leales, aunque notoriamente menor en valor absoluto (–0,0024 + 0,0022 = –0,0002). Este resultado nos indica, por tanto, que los hogares leales son menos sensibles al precio que los no leales y que esta diferencia es significativa. El coeficiente b_2 es también muy significativo y su alto valor positivo muestra que la variable *PREF* tiene gran importancia en la explicación de la elección de los consumidores. Si comparamos el valor y la significatividad de esta variable en las ecuaciones (4) y (5), podemos ver que en la primera, parte del efecto de esta variable es recogido por la variable interactiva *P* 3 *LEAL,* lo que se observa por la disminución del valor y de la significatividad de b_2 en (4) en relación a (5). El estadístico de bondad de ajuste o^2_{aj} (similar al R^2 en el modelo de regresión) de la ecuación (4) es 0,42 y supone un incremento significativo de algo más del 10 por

CUADRO 5.L.6. Estimación de un modelo sin y con efecto interactivo de las variables dicotómicas de lealtad (esc. (4) y (5))

Modelo	*Sólo constantes*		*Sin efecto interactivo (EC. (4))*		*Con efecto interactivo (EC. (5))*	
Variables	*Coeficiente*	*t*	*Coeficiente*	*t*	*Coeficiente*	*t*
Constantes						
Nacionales 4kg	2,03	25,66	1,08	8,40	1,19	8,93
Distrib-Hiper4 kg	–1,13	–7,50	–0,54	–3,22	–0,47	–2,84
Distrib-Super 4kg	–0,03	–0,37	–0,08	–0,65	0,00	0,00
Segundas 2 kg	–0,76	–5,76	–0,31	–1,99	–0,36	–2,37
Nacionales 2 kg	1,94	24,43	1,45	10,80	1,56	11,18
Distrib-Hiper 2 kg	–1,93	–9,22	–1,13	–5,23	–1,20	–5,60
Distrib-Snper 2 kg	–0,20	–1,82	0,26	2,07	0,24	1,87
Preferencia			3,51	48,4	2,14	20,68
Precio			–0,0012	–3,5	–0,0024	–5,96
Precio x dicot. lealtad					0,0022	17,77
N obs.	*3305*		*3305*		*3305*	
$-L$	*4581,45*		*2841,6*		*2680,55*	
ρ^2_{aj}	*0*		*0,38*		*0,42*	

Notas: Para explicación de las medidas de bondad de ajuste, ver nota al pie del Cuadro 5.

[12] La base de datos a nuestra disposición nos proporciona únicamente información del precio de la alternativa comprada en cada ocasión. El precio de las variedades no elegidas se ha estimado como la media de los precios observados de cada variedad durante la semana en que ha tenido lugar la compra. Éste es un procedimiento habitual en la literatura (Krishnamurthi *et al.* 1991).

[13] Esta especificación nos sirve de base para el cálculo de la medida de bondad de ajuste ρ^2_{aj} cuya descripción aparece al pie del Cuadro 5.L.5.

CUADRO 5.L.7. Elasticidades-precio para los consumidores leales y no leales a cada variedad a partir de los resultados de las estimaciones (6) y (7)

Variedad	*Precio medio por paquete*	*Cuota*	*Elasticidad no leales*	*Elasticidad leales*	*Elasticidad todos*
Segundas 4kg	493,1	0,05	–1,12	–0,09	–0,56
Nacionales 4 kg	693,8	0,42	–0,97	–0,08	–0,48
Distrib-Riper 4 kg	549,9	0,02	–1,29	–0,11	–0,65
Distrib-Super4kg	560,4	0,06	–1,26	–0,11	–0,63
Segundas 2 kg	598,6	0,02	–1,41	–0,12	–0,70
Nacionales 2 kg	712~I	0,37	–1,08	–0,09	–0,54
Distrib-Hiper 2 kg	477,7	0,01	–1,14	–0,09	–0,57
Distrib-Super 2 kg	496,2	0,04	–1,14	–0,10	–0,57
Suma		1,00			

ciento sobre el 0,38 de la ecuación (5)[14] Esta superioridad pone de manifiesto la importancia de la segmentación realizada en la explicación del comportamiento de los hogares.

Las elasticidades-precio toman la forma (BenAkiva y Lerman 1987)

$$e_{S_j P_j} = bP_j[I - S_j]$$

donde S_j y P_j son respectivamente la cuota y el precio medio observados de la variedad j.

En el Cuadro 5.L.7 se han calculado las elasticidades-precio de cada una de las variedades para todo el mercado y para el segmento de leales y no leales de forma separada. Se observa que en el segmento de los leales las elasticidades son siempre cercanas a cero (la más alejada es –0,12), mientras que en el de los no leales están por debajo de –1 (excepto para las marcas nacionales de 4 kg que es –0,97). Si consideramos el mercado sin segmentar las elasticidades se sitúan siempre, lógicamente, entre las encontradas en los dos segmentos. También es interesante observar que las más inelásticas son las marcas nacionales, lo que refleja su mayor poder de mercado y el que obtengan las mayores cuotas de mercado teniendo unos precios tan superiores a las demás marcas. Este resultado está en consonancia con su mayor grado de lealtad.

A continuación se presentan las conclusiones mas importantes del trabajo y los comentarios finales.

3. CONCLUSIONES Y COMENTARIOS FINALES

Este trabajo presenta una metodología que puede ser llevada a cabo por las empresas para evaluar la conveniencia de dividir su actividad de marketing entre la adquisición y la retención de clientes, así como para encontrar el balance óptimo en la asignación de recursos entre ambos esfuerzos, como sugieren Blattberg y Deighton (1996) y Sheth y Sisodia (1995). Partiendo del hecho comúnmente aceptado de que es más rentable retener a un cliente leal que a uno no leal, mediante la segmentación del mercado según el grado de lealtad de los consumidores, las empresas pueden centrar sus esfuerzos en la retención de los clientes leales por un lado y en la adquisición de los no leales por otro. La aplicación empírica sobre el mercado de los detergentes durante los años 1992 y 1993 ha permitido identificar y caracterizar a los consumidores leales frente a los no leales. Ambos segmentos son accesibles para las empresas ya que, entre otras características, se diferencian en el tipo de establecimiento comercial preferido y en su tasa de consumo. Asimismo, se ha analizado su respuesta diferencial a las variables de marketing. En concreto se ha estimado la elasticidad-precio de ambos segmentos observándose que la demanda de los consumidores leales es sensiblemente más inelástica que la de los no leales. Con toda esta información la empresa es capaz de evaluar la rentabilidad de planes de marketing alternativos en relación a lo comentado al principio de este pánafo y de tomar decisiones adecuadas al respecto.

Otra de las conclusiones obtenidas es que la segmentación realizada permite aumentar considerablemente (en un 10 por ciento) el grado de explicación del comportamiento de los consumidores. Esto se ha observado comparando la capacidad explicativa de las especificaciones de la démanda (4) y (5) que aparecen en el Cuadro 5.L.6.

Por otro lado se ha encontrado que el grado de lealtad difiere entre las marcas. Las marcas líderes (las nacionales en este caso) no solo tienen más clientes leales sino que también lo son en mayor grado. Asimismo se ha encontrado que estas marcas tienen una demanda más inelástica que el resto, lo que está en consonancia con su mayor grado de lealtad. Pensamos que este caso es un buen ejemplo de la labor del marketing en general y de la publicidad en particular, ya que la principal diferencia entre las marcas líderes y el resto no es su calidad intrínseca[15] sino su considerablemente

[14] El estadístico de contraste es $X^2 = -2[(-L(\beta[3]) - (-L(\beta[6])]$, que se distribuye como una χ^2 con un número de grados de libertad igual a la diferencia de parámetros entre ambos modelos (BenAkiva y Lerman 1985). En este caso $X^2 = -2[(-2.841,6) - (-2.680,55)] = 322,2$ que es significativo a un nivel $p < 0,01$.

[15] Ya hemos mencionado que numerosos análisis concluyen que la calidad de los detergentes es similar en los distintos tipos de marcas.

mayor inversión publicitaria. Esta inversión es capaz de hacer percibir como de mayor calidad a marcas que en realidad la tienen semejante y de convertirlas en marcas líderes vendiéndolas a un precio muy superior.

Otra conclusión interesante proviene del hecho de que son los clientes no leales los que en el mercado analizado adoptan la innovación del nuevo formato de detergente concentrado. Este hecho parece indicar que los consumidores no leales son más innovadores y que, por tanto, las empresas y marcas pueden «adquirirlos» con mayor facilidad.

Una de las características que definen a los no leales es que compran principalmente en super e hipermercados, frente a los leales que compran más en otro tipo de tiendas. Esto sugiere que parte del control de Ta gestión de la lealtad y retención de clientes está en manos de los detallistas y por tanto fuera del control de los fabricantes. Por tanto, las políticas de retención de clientes serán más efectivas si existe coordinación vertical entre ambos agentes. Por otro lado si, como apuntan Stern y ElAnsary (1992) y corrobora para el mercado español el informe del Instituto Nacional del Consumo (1996), las ventas en las grandes superficies van a aumentar y, de manera paralela, el poder de los detallistas va a crecer (Segal-Horn y McGee 1989), parece claro que el papel de los detallistas en la adquisición y retención de clientes será cada vez más importante.

Otra conclusión que se desprende de nuestro análisis se refiere al valor de las marcas analizadas. Entre los criterios que señala Aaker (1994) para medirlo figuran el diferencial de precios valor y las diferencias de cuota de mercado y de lealtad que las marcas de alto valor respecto a las de bajo. Múgica y Yagüe (1993) analizan el valor de las marcas de diversas categorías de productos del mercado de la alimentación en España. Considerando que las marcas líderes en nuestro mercado son las nacionales[16], y siendo conscientes de que estamos hablando de productos diferentes, podemos decir, con las salvedades apuntadas, que en el mercado de los detergentes el valor de las marcas líderes es relativamente alto. Esto puede observarse puesto que las marcas líderes en nuestro caso poseen el 79 por ciento de la cuota de mercado y su precio es casi un 32 por ciento superior al de las demás marcas, cifras elevadas en comparación con los resultados que estos autores hallan para el mercado de la alimentación.

Finalmente, mencionar que este tipo de metodología puede ser implementada y es de utilidad para las empresas. Consideramos que la segmentación del mercado en leales y no leales puede ser de utilidad para analizar comparativamente la rentabilidad de planes de marketing alternativos y para la asignación de recursos entre las estrategias de adquisición y retención de clientes. Asimismo, el conocimiento del índice de lealtad proporciona un valor más exacto de los resultados alcanzados por la actividad de marketing y el valor de las marcas de la empresa.

[16] Múgica y Yagüe mencionan que el grado de validez de estos análisis a nivel empírico depende en gran medida de la precisión con que sea delimitado el grupo de las marcas líderes.

BIBLIOGRAFÍA

AAKER, D. A. (1994). *Gestión del valor de marca.* Madrid: Díaz de Santos.

BALDINGER, A. (1992). «What CEOs are Saying About Braud Equity», *Jaurnal of Advertising Research,* Vol. 34 (4), pp. RC6-RCI3.

BEN-AKIVA, M., y LERMAN, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge.* MIT Press, Mass.

BERNÉ, C. (1994). «El Enfoque del Marketing en la Productividad: una Aplicación al Comercio Minorista», *Tesis Doctoral.* Universidad de Zaragoza.

BLACKSTONE, M. (1990). «Price Trade-Offs as a Measure of Brand Value», *Journal of Advertising Research,* Vol. 3 (4), pp. RC3-RC6.

BLATTBERG, R. C. y DEIGHTON, J. (1996). «Manage Marketing by the Consumer Equity», *Harvard Business Review* (July-August), pp. 136-144.

CEBOLLADA, J. (1996). «Lealtad de Marca, Competencia y Estrategia de Marketing. Aplicación a un Producto no Duradero», *Tesis Doctoral.* Barcelona: Universidad Pompeu Fabra.

CIUDADANO (1992). «El Blanco más Blanco», (Octubre), Madrid: Unión de Consumidores de España.

CIUDADANO (1994). «¿Quién Lava más Blanco?», (Noviembre, Madrid: Unión de Consumidores de España.

COLOMBO, R. A. y MORRISON, D. G. (1983). «A Brand Switching Model with Implications for Marketing Strategies», *Marketing Science,* Vol. 8 (1), pp. 89-99.

EROSKI (1994). «Detergentes Concentrados y Normales», *Eroski Mondragon,* 187 (Septiembre).

FORNELL, C. y WERNERFELT, B. (1987). «Defensive Marketing Strategy by Customer Complaint Management», *Journal of Marketing Research,* 24 (November), pp. 337-346.

FRANK, F. M. (1967). «Is Brand Loyalty a Useful Basis for Segmentation?», *Journal of Advertising Research* (June).

GREENACRE, M. J. (1993). *Correspondence Analysis in Practice.* England: London Academic Press.

GROVER, R. y SRINIVASAN, V. (1987). «A Simultaneous Approach to Market Segmentation and Market Structuring», *Journal of Marketing Research,* 29 (Nov), pp. 474-476.

GUADAGNI, P. M. y LITTLE, J. D. C. (1983), «A Logit Model of Brand Choice Calibrated on Scanner Data», *Marketing Science,* Vol. 2 (3), pp. 203-238.

HERNÁNDEZ, M., MUNUERA, J. L. y RUÍZ, S. (1995). «La Estrategia de Diferenciación en el Comercio Minorista», *Información Comercial Española,* 739 (Marzo), pp. 27-45.

INSTITUTO NACIONAL DEL CONSUMO (1996). *El consumidor español: presente y futuro.* Madrid: Ministerio de Sanidad y Consumo.

JACOBY, J. y CHESNUT, R. W. (1978). *Brand Loyalty: Measurement and Management.* Nueva York: Wiley & Sons.

KLEMPERER, P. (1987). «Markets with Consumer Switching Costs», *Quarterly Journal of Economics,* 102 (May).

KRISHNAMURTHI, L. Y RAJ, S. P. (1991). «An Empirical Analysis of the Relationship Between Brand Loyalty and Consumer Price Elasticity», *Marketing Science,* Vol. 10 (2), pp. 172-183.

MASSY, W. F. Y FRANK, F. M. (1965). «Short-tenn Price and Dealing Effects in Selected Market Segments», *Journal of Marketing Research,* 2 (May).

MCCANN, J. M. (1974). «Marketing Segment Response to the Marketing Decision Variables», *Journal of Marketing Research,* 11 (November), pp. 399-412.

MÚGICA, J. M. Y YAGÜE, M. J. (1993). «Impacto del Capital Comercial en la Competitividad Empresarial», *Papeles de Economía Española,* 56, pp. 242-252.

NESLIN, A. S., HENDERSON, C. Y QUELCH, J. (1985). «Consumer Promotions and the Acceleration of Product Purchases», *Marketing Science,* Vol. 4 (2) pp. 147-165.

NUENO, J. L. Y Ros, P. (1996). «La Retención de Clientes en Distribución», *Marketing y Ventas.* Dossier: Las Nuevas Estrategia de la Distribución en España.

OCU-COMPRA MAESTRA (1995). «Detergentes en Polvo Concentrados», (Marzo), Madrid: Organización de Consumidores y Usuarios.

OLSON, J. C. Y JACOBY, J. (1971). «A Construct Validation Study of Brand Loyalty», *Proceedings* Vol. 6, American Psychological Association, pp. 657-658.

REICHHELD, E. E. Y SASSER, W. E. (1990). «Zero Defections: Quality Comes to Service», *Harvard Business Review,* 68 (Oct./Nov.) pp. 105-111.

SARABIA, E. J. Y MUNUERA, J. L. (1994). «Concepto y Usos de la Segmentación de Mercados: Una Perspectiva Teórica y Práctica», *Información Comercial Española,* 727 (Marzo), pp. 111-124.

SEGAL-HORN, S. Y MCGEE, J. (1989). «Strategies to Cope with Retailing Bying Power», en Pellegriny, L. y Reddy, S. K. (eds.), *Retail and Marketing Channels.* Londres: Routledge.

SHETH, J. N. Y SISODIA, R. S. (1995). «Feeling the Heat», Partes I y II, *Marketing Management,* Vol. 4 (2), pp.8-23 y Vol. 4 (2), pp.19-33.

STARR, M. K. Y RUBINSON, J. R. (1978). «A Loyalty Group Segmentation Model for Brand Purchasing Simulation», *Journal of Marketing Research,* 15 (August), pp. 378-383.

STERN, L. W. Y EL-ANSARY, A. I. (1992). *Marketing Channels.* 4th Ed., Englewood Cliffs: Prentice Hall Int.

CAPÍTULO 6

Análisis multivariante de la varianza

OBJETIVOS DE APRENDIZAJE

- Explicar la diferencia entre la hipótesis nula univariante del ANOVA y la hipótesis nula multivariante del MANOVA.
- Discutir las ventajas del enfoque multivariante para contrastar la significación en comparación con los enfoques univariantes más tradicionales.
- Establecer los supuestos básicos para el uso del MANOVA.
- Discutir los diferentes tipos de contrastes estadísticos que están disponibles para probar la significación en el MANOVA.
- Describir los objetivos de los contrastes post hoc en el ANOVA y en el MANOVA.
- Interpretar los resultados de la interacción cuando se emplea en más de una variable independiente el MANOVA.
- Describir los objetivos del análisis multivariante de la covarianza (MANCOVA).

Como construcción teórica, el análisis multivariante de la varianza (MANOVA) fue introducido hace varias décadas según la formulación original de Wilks [22]. Sin embargo, hasta que no se desarrollaron contrastes apropiados con distribuciones tabuladas y se tuvo una amplia disponibilidad de programas para calcular estos contrastes en computadores con alta velocidad, el MANOVA no llegó a ser una herramienta práctica para los investigadores.

El análisis multivariante de la varianza es una extensión del análisis de la varianza (ANOVA) en donde se tiene en cuenta más de una variable de criterio. Es una técnica de dependencia que mide las diferencias de dos o más variables métricas dependientes basadas en un conjunto de variables categóricas (no métricas) que actúan como predictores. El MANOVA y el ANOVA pueden ser representados según las siguientes expresiones generales:

Análisis de la varianza

$$\underset{\text{(métricas)}}{Y_1 + Y_2 + Y_3 + \ldots + Y_n} = \underset{\text{(no métricas)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

Análisis multivariante de la varianza

$$\underset{\text{(métrica)}}{Y_1} = \underset{\text{(no métricas)}}{X_1 + X_2 + X_3 + \cdots + X_n}$$

Al igual que el ANOVA, el MANOVA se refiere a diferencias entre grupos (o tratamientos de experimentos). Sin embargo, el ANOVA se denomina como proceso univariante debido a que se emplea para valorar las diferencias entre grupos utilizando una única variable dependiente métrica. El MANOVA se denomina como proceso multivariante debido a que se usa para valorar diferencias entre grupos a través de múltiples variables dependientes métricas de forma simultánea. En el MANOVA, cada grupo de tratamiento es definido por dos o más variables dependientes.

Tanto el ANOVA como el MANOVA son particularmente útiles cuando se utilizan conjuntamente con **diseños de experimentos** —esto es, los diseños de investigación en los que el investigador controla o manipula directamente una o más variables independientes para determinar el efecto sobre una (ANOVA) o más (MANOVA) variables dependientes. El ANOVA y el MANOVA proporcionan las herramientas necesarias para juzgar la fiabilidad de cualquier efecto observado (por ejemplo, si una diferencia observada es debida a un efecto del tratamiento o a la variabilidad aleatoria del muestreo). Véase en el Capítulo 1 una discusión de cómo estas técnicas se relacionan con otros procedimientos univariantes.

¿Qué es el análisis multivariante de la varianza?

El análisis multivariante de la varianza es la extensión multivariante de las técnicas univariantes para valorar las diferencias entre las medias de los grupos. El procedimiento univariante utiliza el contraste *t* para situaciones con dos grupos, y el análisis de la varianza (ANOVA) en situaciones con tres o más grupos definidos por dos o más variables independientes. Antes de proceder con nuestras discusiones sobre los aspectos propios del análisis multivariante de la varianza, repasaremos los principios básicos de las técnicas univariantes.

Procedimientos univariantes para valorar las diferencias de grupo

Se clasifican estos procedimientos como univariantes no por el número de variables independientes, sino por el número de variables dependientes. En la regresión múltiple, los términos univariante y multivariante se refieren al número de variables independientes, pero para ANOVA y MANOVA, se aplica la terminología al uso de variables simples o múltiples. El debate que se detalla a continuación aborda los dos tipos de procedimientos univariantes más comunes, el contraste *t*, que compara una variable dependiente a través de dos grupos y ANOVA, que se utiliza cuando el número de grupos es tres o más.

El contraste t

El **contraste *t*** valora la significación estadística de las diferencias entre dos medias muestrales independientes. Por ejemplo, un investigador puede considerar dos grupos de encuestados para

diferentes anuncios publicitarios que reflejan diferentes mensajes, uno informativo y el otro emocional, y consecuentemente pregunta a cada grupo sobre la pretensión del mensaje en una escala de diez puntos, donde uno significa pobre y diez significa excelente. Los dos mensajes publicitarios distintos representan un **tratamiento** con dos niveles (informativo frente a emocional). Un tratamiento también conocido como un **factor**, es una variable independiente no métrica, manipulada u observada experimentalmente, que puede ser representada en varios categorías o niveles. En nuestro ejemplo, el tratamiento es el efecto de la pretensión emocional frente a la informativa.

Para determinar si los dos mensajes son percibidos de manera diferente (significando que el tratamiento tiene un efecto), se calcula un estadístico *t*. El **estadístico *t*** es el ratio de las diferencias de las medias muestrales ($\mu_1 - \mu_2$) y su error estándar. El **error estándar** es una estimación de la diferencia entre las medias que se espera debido al error muestral, más que debido a diferencias reales entre las medias. Se puede mostrar esto con la ecuación

$$\text{Estadístico } t = \frac{\mu_1 - \mu_2}{SE_{\mu_1\mu_2}}$$

donde

μ_1 = media del grupo 1

μ_2 = media del grupo 2

$SE_{\mu_1\mu_1}$ = error estándar de la diferencia en las medias de grupo

Al construir el ratio entre la diferencia real entre las medias y la diferencia esperada debido al error muestral, cuantificamos la cantidad de impacto real del tratamiento que se debe al error muestral aleatorio. Dicho de otra manera, el valor *t* o estadístico *t*, representa la diferencia de grupo en términos de errores estándar.

Si el valor *t* es suficientemente grande, entonces podemos decir, estadísticamente hablando, que la diferencia no se debe a la variabilidad del muestreo, sino que representa una diferencia real. Esto se lleva a cabo con la comparación del estadístico *t* con el **valor crítico** del estadístico (t_{crit}). Si el valor absoluto del estadístico *t* excede al valor crítico, esto lleva al rechazo de la **hipótesis nula** de que no existen diferencias en las pretensiones de los mensajes publicitarios entre los grupos, lo que quiere decir que la diferencia real debida a las pretensiones es mayor que la diferencia esperada a partir del error en el muestreo.

Determinamos un valor crítico (t_{crit}) para nuestro estadístico *t* y contrastamos la significación estadística de las diferencias observadas de la siguiente forma:

1. Calcular el estadístico *t* como el ratio de la diferencia entre la media de la muestra y su error estándar.
2. Especificar un nivel de **error de Tipo I** (denotado por α o **nivel de significación**), que indica el nivel de probabilidad que el investigador aceptará para concluir que las medias de los grupos son diferentes cuando realmente no lo son.
3. Determinar el valor crítico (t_{crit}) mediante la remisión a la distribución *t* con $N_1 + N_2 - 2$ grados de libertad y con el α especificado, donde N_1 y N_2 representan los tamaños muestrales.
4. Si el valor absoluto del estadístico *t* calculado excede al t_{crit}, el investigador puede concluir que los dos mensajes publicitarios tienen diferentes niveles de pretensión (por ejemplo, $\mu_1 \neq \mu_2$), con una probabilidad de error de Tipo I de α. El investigador puede entonces examinar los valores medios reales para determinar qué grupo es mayor en el valor dependiente.

Análisis de la varianza

En nuestro ejemplo sobre el contraste *t*, el investigador consideró dos grupos de encuestados para diferenciar los mensajes publicitarios, y para ello se les pidió que clasificaran las pretensiones de los anunciantes en una escala de diez puntos. Supongamos que estuviéramos interesados en eva-

luar tres mensajes publicitarios en vez de dos. Los encuestados serían asignados aleatoriamente a uno de los tres grupos, y tendríamos tres medias muestrales para comparar. Para analizar estos datos, podíamos estar tentados a llevar a cabo contrastes *t* separados para contrastar la diferencia entre cada par de medias (por ejemplo, grupo 1 frente a grupo 2; grupo 1 frente a grupo 3; y grupo 2 frente a grupo 3).

Sin embargo, los contrastes *t* múltiples hinchan el porcentaje del error Tipo I global (abordaremos esto con más detalle en la siguiente sección). El **ANOVA** evita este aumento del error de Tipo I al comparar un conjunto de grupos de tratamiento, determinando si el conjunto completo de medias muestrales indica que las muestras fueron tomadas de la misma población general. Es decir, el ANOVA es empleado para determinar la probabilidad de que las diferencias en las medias entre varios grupos sean debidas meramente al error muestral.

La lógica de un contraste ANOVA es bastante simple. Como indica el nombre «análisis de la varianza», se comparan dos cálculos independientes de la varianza para la variable independiente: uno que refleja la variabilidad general de los encuestados entre los grupos (CM_I) y otro que representa las diferencias entre los grupos que se atribuyen a los efectos del tratamiento (CM_E):

1. *El cálculo de la varianza dentro de los grupos* (CM_I: cuadrado medio intra grupos): es una estimación de la variabilidad aleatoria de los encuestados sobre la variable dependiente dentro de un grupo de tratamiento, y está basado en desviaciones de puntuaciones individuales respecto de las medias de sus grupos respectivas, pero no incluye las diferencias entre las medias de los grupos. El cuadrado medio intra grupos es comparable al error estándar entre dos medias calculado en el contraste *t* dado que representa la variabilidad dentro de los grupos. Se denomina también varianza del error.
2. *El cálculo de la varianza entre los grupos* (CM_E: cuadrado medio entre grupos): la segunda estimación de la varianza es la variabilidad de las medias de los grupos de tratamiento sobre la variable dependiente. Se basa en desviaciones de las *medias de los grupos* respecto de la media global de todas las puntuaciones. Bajo la hipótesis nula de que no hay efectos de tratamiento (por ejemplo, $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$), esta varianza, al contrario que el cuadrado medio intra grupos, refleja cualquier efecto de tratamiento que exista, es decir, las diferencias en el tratamiento implican un incremento en el valor esperado del cuadrado medio entre grupos.

Dado que la hipótesis nula de no existencia de diferencias entre los grupos es cierta, los cuadrados medios intra y entre grupos representan estimaciones independientes de la varianza poblacional. Por tanto, su ratio es una medida de cuánta varianza es atribuible a los diferentes tratamientos frente a la varianza esperada del muestreo aleatorio. Este ratio nos proporciona un valor de un estadístico *F*, cuyo cálculo se parece al del valor *t*, y se podría mostrar como

$$\text{Estadístico } F = \frac{CM_E}{CM_I}$$

Dado que las diferencias entre los grupos tienden a aumentar el CM_E, valores grandes del estadístico *F* nos llevan al rechazo de la hipótesis nula de que no existen diferencias en las medias de los grupos. Si el análisis tiene varios tratamientos diferentes (variables independientes), entonces los valores del cuadrado medio entre grupos se calculan para cada tratamiento, con lo que se permite la valoración separada para cada tratamiento.

Para determinar si el estadístico *F* es suficientemente grande para apoyar el rechazo de la hipótesis nula, se sigue un proceso similar al del contraste *t*. Primero, determinar el valor crítico para el estadístico *F* (F_{crit}) atendiendo a la distribución *F* con $(k - 1)$ y $(N - k)$ grados de libertad para un nivel dado (donde $N = N_1 + \ldots + N_k$ y k = número de grupos). Si el valor del estadístico *F* calculado es mayor que F_{crit} concluiremos que las medias entre los grupos no son iguales.

El examen de las medias de los grupos permite entonces valorar la importancia relativa de cada grupo sobre la medida dependiente. Aunque el contraste *F* contrasta la hipótesis nula de igualdad de las medias, no resuelve la cuestión de qué medias son diferentes. Por ejemplo, en una situación

con tres grupos, los tres grupos pueden diferir significativamente, o dos pueden ser iguales pero diferir del tercero. Para valorar estas diferencias, el investigador puede emplear comparaciones planificadas o contrastes post hoc. Examinamos cada uno de estos métodos en una sección posterior.

Análisis multivariante de la varianza (MANOVA)

Como procedimientos de inferencia estadística, tanto las técnicas univariantes (contraste *t* y ANOVA) como el MANOVA se emplean para contrastar la significación estadística de las diferencias entre los grupos. En el contraste *t* y en el ANOVA, la hipótesis nula contrastada es la igualdad de las medias de las variables dependientes entre los grupos. En el MANOVA, la hipótesis nula contrastada es la igualdad de **vectores** de medias de variables dependientes múltiples entre los grupos. La diferencia entre las hipótesis contrastadas en el ANOVA y en el MANOVA está representada en la Figura 6.1. En el caso univariante se contrasta la igualdad entre los grupos de una única variable dependiente. En el caso multivariante, se contrasta la igualdad de un **valor teórico**. El concepto de un valor teórico ha sido fundamental en discusiones previas sobre las técnicas multivariantes y se trató con detalle en el Capítulo 1. En el MANOVA, el investigador tiene realmente dos valores teóricos, uno construido a partir de las variables dependientes y el otro a partir las variables independientes. El valor teórico con las variables dependientes es más interesante, ya que las medidas dependientes métricas pueden ser introducidas en una combinación lineal como ya se vio en la regresión múltiple y en el análisis discriminante. El aspecto propio del MANOVA es que el valor teórico combina óptimamente las medidas dependientes múltiples dentro de un valor único que maximiza las diferencias entre los grupos.

ANOVA

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$

Hipótesis nula (H_0) = las medias de todos los grupos son iguales, es decir, provienen de la misma población.

MANOVA

$$H_0 : \begin{matrix} \mu_{11} \\ \mu_{21} \\ \\ \\ \mu_{p1} \end{matrix} = \begin{matrix} \mu_{12} \\ \mu_{22} \\ \\ \\ \mu_{p2} \end{matrix} = \ldots\ldots = \begin{matrix} \mu_{1k} \\ \mu_{2k} \\ \\ \\ \mu_{pk} \end{matrix}$$

Hipótesis nula (H_0) = los vectores de las medias de todos los grupos son iguales, es decir, provienen de la misma población.

μ_{pk} = Media de la variable *p*, del grupo *k*

FIGURA 6.1. Contraste de la hipótesis nula del ANOVA y del MANOVA.

El caso de dos grupos: T^2 de Hotelling

En nuestro ejemplo univariante anterior, los investigadores estaban interesados en la pretensión de dos mensajes publicitarios. Pero, ¿qué ocurre si también quisieran conocer la intención de compra generada por los dos mensajes? Si se emplease solamente análisis univariante, los investigadores llevarían a cabo contrastes *t* separados tanto para la pretensión de los mensajes como para

la intención de compra generada por los mensajes. Sin embargo, las dos medidas no están relacionadas; por ello, lo deseable es un contraste de las diferencias entre los mensajes en ambas variables conjuntamente. Aquí es donde se pueden usar la **T^2 de Hotelling**, una forma especializada del MANOVA que es una extensión directa del contraste *t* univariante.

La **T^2 de Hotelling** proporciona un contraste estadístico del valor teórico formado por las variables dependientes que produce la mayor diferencia entre los grupos. También tiene en cuenta el problema de «inflar» la tasa del error de Tipo I que aumenta al realizar una serie de contrastes *t* para la comparación entre las medias de los grupos sobre las medidas dependientes. Este contraste controla el aumento de la tasa del error de Tipo I proporcionando un único contraste general para contrastar las diferencias de los grupos entre todas las variables dependientes para un nivel de significación α dado.

¿Cómo logra la T^2 de Hotelling estos objetivos? Considere la siguiente ecuación de un valor teórico de las variables dependientes:

$$C = W_1Y_1 + W_2Y_2 + \ldots + W_nY_n$$

donde

C = puntuación del valor teórico o de la combinación para un encuestado

W_i = ponderación para la variable dependiente *i*

Y_i = variable dependiente *i*

En nuestro ejemplo, las clasificaciones de las pretensiones de los mensajes son combinadas con las intenciones de compra para formar la combinación. Para cualquier conjunto de ponderaciones, podríamos calcular puntuaciones de la combinación para cada encuestado, y después calcular un estadístico *t* normal para la diferencia entre los grupos sobre las puntuaciones de la combinación. Sin embargo, si nosotros encontramos un conjunto de ponderaciones que generan el valor máximo del estadístico *t* para este conjunto de datos, estas ponderaciones serían las mismas que las de la función discriminante entre los dos grupos (lo cual no debe sorprender a aquellos que han leído el Capítulo 5). El estadístico *t* máximo que se obtiene con las puntuaciones de la combinación elaboradas a partir de la función discriminante pueden ser elevadas al cuadrado para obtener el valor de la T^2 de Hotelling [10]. La expresión para el cálculo de la T^2 de Hotelling representa los resultados de las derivaciones matemáticas empleadas para resolver un estadístico *t* máximo (e implícitamente, la combinación lineal más discriminante de las variables dependientes). Esto es equivalente a decir que si nosotros podemos encontrar una función discriminante para los grupos que producen una T^2 significativa, los dos grupos son considerados diferentes entre los vectores de las medias.

¿Cómo proporciona la T^2 un contraste para la hipótesis de que no existen diferencias de los grupos sobre los vectores de puntuaciones medias? Al igual que el estadístico *t* sigue una distribución conocida bajo la hipótesis nula de que no existen efectos del tratamiento sobre una única variable dependiente, la T^2 de Hotelling sigue una distribución conocida bajo la hipótesis nula de que no existen efectos del tratamiento sobre cualquiera de los conjuntos de las medidas dependientes. Esta distribución es una distribución F con p y $N_1 + N_2 - 2 - 1$ grados de libertad del ajuste (donde p = el número de variables dependientes). Para obtener el valor crítico para la T^2 de Hotelling, encontramos el valor tabulado para F_{crit} a un nivel de significación α dado y se calcula T^2_{crit} como sigue:

$$T^2_{crit} = \frac{p(N_1 + N_2 - 2)}{N_1 + N_2 - p - 1} \times F_{crit}$$

El caso de k *grupos: MANOVA*

El MANOVA puede ser considerado como una sencilla extensión del procedimiento de la T^2 de Hotelling; es decir, nosotros ideamos ponderaciones de las variables dependientes para elaborar una puntuación del valor teórico para cada encuestado, como se describió anteriormente. Si quisiéramos evaluar tres mensajes publicitarios tanto en sus pretensiones como en las intenciones de com-

pra que ellos generan, emplearíamos el MANOVA. Con el MANOVA queremos encontrar ahora el conjunto de ponderaciones que maximizan el valor de la F del ANOVA calculado sobre las puntuaciones de los valores teóricos para todos los grupos. Pero el MANOVA también puede ser considerado como una extensión del análisis discriminante (véase Capítulo 5) en donde pueden ser construidos los valores teóricos de las medidas dependientes si el número de grupos es tres o más. El primer valor teórico, denotado como **función discriminante**, establece un conjunto de ponderaciones que maximizan las diferencias entre los grupos, y por tanto maximiza el valor de la F. El valor de la F máximo nos permite calcular directamente lo que es conocido como el **estadístico de la mayor raíz característica (mrc),** que da cabida para el contraste estadístico de la primera función discriminante. El estadístico de la mayor raíz característica puede ser calculado como: $\text{mrc} = (k - 1) F_{\text{max}} / (N - k)$ [11].

Para obtener un único contraste de la hipótesis de que no existen diferencias de los grupos sobre los vectores de las puntuaciones medias, podemos emplear las tablas de la distribución de la mrc. Al igual que el estadístico F seguía una distribución conocida bajo la hipótesis nula de que las medias de los grupos son equivalentes sobre una variable dependiente única, el estadístico mrc sigue una distribución conocida bajo la hipótesis nula de que los vectores de las medias de los grupos son equivalentes (por ejemplo, las medias de los grupos son equivalentes sobre un conjunto de medidas dependientes). La comparación de la mrc observada con la mrc_{crit} nos da una base para rechazar la hipótesis nula global de que los vectores de las medias de los grupos son equivalentes.

Cualquiera de las funciones discriminantes posteriores son **ortogonales**; éstas maximizan las diferencias entre los grupos basándose en la varianza que permanece no explicada por la(s) función(es) anterior(es). Por ello, en muchos ejemplos, el contraste para las diferencias entre los grupos incluye no sólo un único valor teórico sino un conjunto de puntuaciones de valores teóricos que son evaluadas simultáneamente. Se hallan disponibles un conjunto de contrastes multivariantes, cada uno de los cuales es el más apropiado para unas situaciones específicas para la contrastación de estos valores teóricos múltiples.

Las diferencias entre MANOVA y el análisis discriminante

En la sección anterior, discutimos los elementos básicos tanto de los contrastes univariantes como los multivariantes para valorar las diferencias entre los grupos en una o más variables dependientes. De esta manera, observamos el cálculo de la función discriminante; en el caso de MANOVA es el valor teórico de las variables dependientes que maximiza la diferencia entre grupos. La cuestión que puede surgir: ¿cuál es la diferencia entre MANOVA y el análisis discriminante? En algunos aspectos, MANOVA y el análisis discriminante son «imágenes de espejo». Las variables dependientes en MANOVA (una serie de variables métricas) son las variables independientes en el análisis discriminante y una simple variable dependiente no métrica del análisis discriminante se convierte en la variable independiente en el MANOVA. Además, ambos utilizan los mismos métodos en la formación de valores teóricos y evalúan la significación estadística entre los grupos.

Las diferencias, sin embargo, se centran alrededor de los objetivos de los análisis y el papel de la variable no métrica. El análisis discriminante emplea una variable no métrica simple como variable dependiente. Se supone que las categorías de la variable dependiente están dadas y que se utilizan las variables independientes para formar valores teóricos que son diferentes de manera máxima entre los grupos formados por las categorías de la variable dependiente. En el MANOVA, la serie de variables métricas actúan ahora como variables dependientes y el objetivo es encontrar grupos de encuestados que exhiben diferencias sobre la serie de variables dependientes. Los grupos de encuestados no son especificados previamente; en su lugar, el investigador utiliza una o más variables independientes (variables no métricas) para formar grupos. El MANOVA, incluso mientras forma estos grupos, retiene todavía la capacidad de valorar el impacto de cada variable no métrica por separado.

Una ilustración hipotética del MANOVA

Supongamos que un simple ejemplo puede ilustrar los beneficios de la utilización del MANOVA. Supongamos que los investigadores identificaron dos variables no métricas (tipo de producto y status de cliente), que pensaron causarían diferencias en la manera en que la gente evalúa la publicidad para los productos de HATCO. Cada variable tiene dos categorías: tipo de producto (producto 1 frente a producto 2) y el estatus del cliente (cliente actual frente a cliente anterior). Con la combinación de estas dos variables, obtenemos cuatro grupos destacados (producto 1/ cliente anterior, producto 2/cliente anterior, producto 1/cliente, producto 2/cliente). Se puede pedir a los clientes seleccionados en cada grupo valorar los anuncios de HATCO con una escala de 10 puntos en cuanto a su capacidad de atraer la atención y su capacidad de persuasión para comprar (véase Tabla 6.1).

Podemos usar el MANOVA para combinar las dos medidas dependientes (recuerdo y compra) en un valor teórico simple. Supongamos para este ejemplo que las dos medidas dependientes tienen igual ponderación cuando se añaden al valor del valor teórico (total), de forma idéntica al análisis discriminante. Este hecho proporciona un valor de compuesto único con las variables ponderadas para alcanzar unas diferencias máximas entre los grupos. El MANOVA es diferente del análisis discriminante en la manera en que se forman y se analizan los grupos. En este caso, sólo se podría llevar a cabo el análisis discriminante sobre la serie de cuatro grupos, sin distinción de las características del grupo (tipo de producto o estatus del cliente). El investigador podría determinar si el valor teórico tenía una diferencia sustancial a través de los grupos; sin embargo no podría valorar cuáles de las características de los grupos tenían relación con estos grupos. Con MANOVA, sin embargo, el investigador analiza las diferencias en los grupos a la vez que valora si las diferencias se deben a tipo de producto, tipo de cliente o ambos. Por tanto, el MANOVA se centra en el análisis sobre la composición de los grupos basada en sus características (las variables independientes). Esto permite al investigador proponer cualquier número de variables no métricas independientes (dentro de los límites para formar grupos y después buscar diferencias significativas en el valor teórico de la variable dependiente con las variables no métricas específicas).

En la Tabla 6.1, los cuatro grupos tienen medias bastante diferentes sobre la variable compuesta «total» (es decir, 4,25, 8,25, 11,75 y 14,0). La Figura 6.2 indica las cuatro medias de grupo. Las dos líneas conectan los grupos (cliente anterior y cliente actual) para el producto 1 y el producto 2. El análisis discriminante también determinaría si hay diferencias considerables sobre el valor teórico compuesto y tambien que ambas variables dependientes (recuerdo y compra) no contribuyen a las diferencias. MANOVA, sin embargo, va más allá del análisis de las diferencias a través

TABLA 6.1. Ejemplo hipotético del MANOVA

	Producto 1 $\bar{x}_{recuerdo} = 3{,}50$ $\bar{x}_{compra} = 4{,}50$ $\bar{x}_{total} = 8{,}00$				*Producto* 2 $\bar{x}_{recuerdo} = 5{,}50$ $\bar{x}_{compra} = 5{,}625$ $\bar{x}_{total} = 11{,}125$			
Tipo de cliente/línea de producto	*ID*	*Atención*	*Compra*	*Total*	*ID*	*Atención*	*Compra*	*Total*
Cliente anterior	1	1	3	4	5	3	4	7
$\bar{x}_{recuerdo} = 3{,}00$	2	2	1	4	6	4	3	7
$\bar{x}_{compra} = 3{,}25$	3	2	3	5	7	4	5	9
$\bar{x}_{total} = 6{,}25$	4	3	2	5	8	5	5	10
Promedio		2,0	2,25	4,25		4,0	4,25	8,25
Cliente	9	4	7	11	13	6	7	13
$\bar{x}_{recuerdo} = 6{,}00$	10	5	6	11	14	7	8	15
$\bar{x}_{compra} = 6{,}875$	11	5	7	12	15	7	7	14
$\bar{x}_{total} = 12{,}875$	12	6	7	13	16	8	6	14
Promedio		5,0	6,75	11,75		7,0	7,0	14,0

Los valores son respuestas en una escala de diez puntos (1 = bajo, 10 = alto).

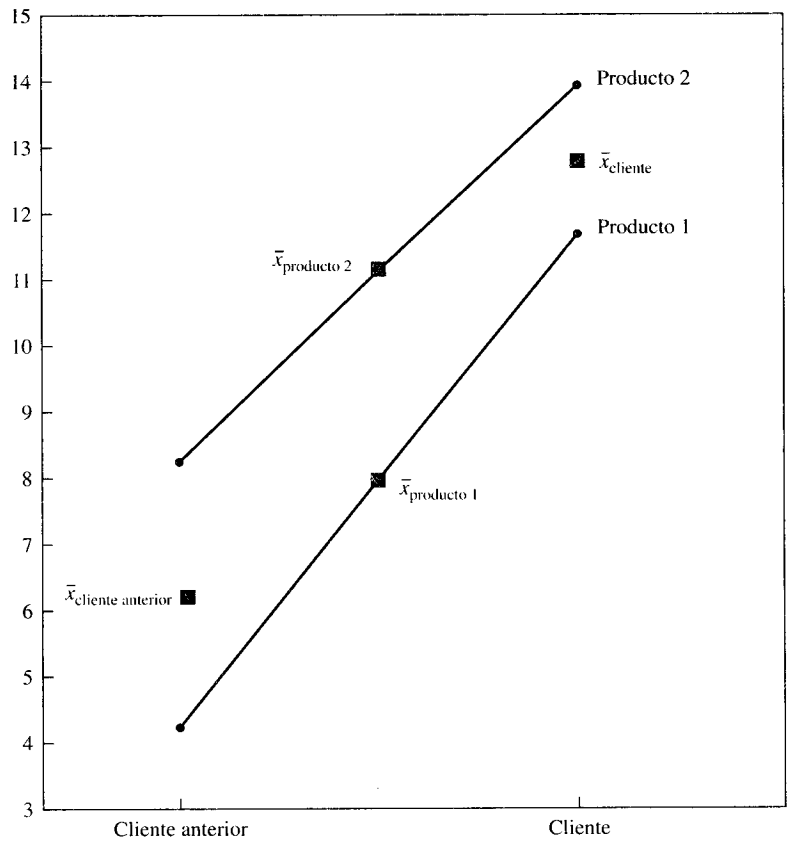


FIGURA 6.2. Representación gráfica de la media del grupo del valor teórico (total) para el ejemplo hipotético.

de los grupos con la evaluación de si el tipo de producto y/o el tipo de cliente crearon grupos con estas diferencias. Esto se consigue con la evaluación de la media de la categoría (indicado por el símbolo ■), que se muestra también en la Figura 6.2. Si observamos el tipo de producto (ignorando las distinciones por lo que se refiere al tipo de cliente), podemos observar un valor medio de 8,0 para los usuarios del producto 1 y un valor medio de 11,125 para los usuarios del producto 2. De la misma manera, para el tipo de cliente, los clientes anteriores tenían un valor medio de 6,25 y los clientes actuales un valor medio de 12,875. Un examen visual sugeriría que las dos medias de categoría demuestran diferencias sustanciales, siendo las diferencias para el tipo de cliente (12,875 – 6,25 = 6,625) más grandes que para el producto (11,125 – 8,00 = 3,125). El MANOVA puede servir para representar estas medias de categoría de la variable independiente en el análisis; no sólo se demuestra que las diferencias globales ocurren (tal y como se hizo con el análisis discriminante) sino que tanto el tipo de cliente como el tipo de producto contribuyen de forma significativa a la formación de estos grupos distintos. Por tanto, ambas características «causan» diferencias significativas, un resultado que no es posible obtener con el análisis discriminante.

¿Cuándo se debe utilizar MANOVA?

El investigador puede obtener diferentes ventajas con el empleo del MANOVA, ya que esta técnica permite el examen de varias medidas dependientes simultáneamente. Trataremos las cuestiones sobre el empleo del MANOVA desde las perspectivas de eficiencia y de control de la precisión estadística mientras nos proporcione la manera apropiada para contrastar cuestiones multivariantes.

Control del porcentaje de errores experimentales

El uso de ANOVAs univariantes separados o de contrastes *t* puede crear un problema cuando se intenta controlar la tasa de errores experimentales o globales [11]. Por ejemplo, suponga que analizamos una serie de cinco variables dependientes utilizando ANOVAs separados, con cada uno se usa 0,05 como nivel de significación. Si no existen diferencias reales entre las variables dependientes, esperaríamos que no se observase ningún efecto de alguna variable dependiente el 5 por ciento de las veces. Sin embargo, en los cinco contrastes separados, la probabilidad del error de Tipo I se situaría entre el 5 por ciento (si todas las variables dependientes están perfectamente correlacionadas) y $(1 - 0{,}95^5)$ 23 por ciento si todas las variables dependientes no están correlacionadas. Por ello un conjunto de contrastes separados no nos permite ningún control de nuestro porcentaje efectivo del error de Tipo I. Si el investigador quiere mantener el control sobre el porcentaje del error experimental y existe algún grado de correlación entre las variables dependientes, entonces la técnica más apropiada es el MANOVA.

Diferencias entre una combinación de variables dependientes

Una serie de contrastes univariantes ANOVA tampoco tiene en cuenta la posibilidad de que alguna combinación (combinación lineal) de las variables dependientes pueda proporcionar evidencia de la existencia de alguna diferencia global en los grupos, que puede no detectarse si se examina cada variable dependiente separadamente. Los contrastes individuales ignoran las correlaciones entre las variables dependientes y por ello no se emplea toda la información disponible para valorar diferencias globales en los grupos. Si existe multicolinealidad entre las variables dependientes, el MANOVA será más potente que las contrastes univariantes separados. De esta manera, el MANOVA puede detectar diferencias *combinadas* que no se encuentran con los contrastes univariantes. Además, si se construyen valores teóricos múltiples, entonces estos pueden proporcionar *dimensiones* de diferencias que puedan distinguir entre los grupos mejor que las variables de forma independiente. Sin embargo, en algunos casos donde existe un gran número de variables dependientes, la potencia de los contrastes ANOVA excede lo que se obtiene empleando un único MANOVA [4]. Las cuestiones relacionadas con el tamaño muestral, número de variables dependientes y la potencia estadística serán consideradas en una sección posterior.

Un proceso de decisión para el MANOVA

El proceso para llevar a cabo el análisis multivariante de la varianza es similar al existente en otras muchas técnicas multivariantes, por lo que puede ser descrito a través del proceso de construcción de un modelo en seis pasos descrito en el Capítulo 1. Este proceso comienza con la especificación de los objetivos de la investigación. Después continua con un conjunto de cuestiones de diseño a las que se enfrente el análisis multivariante y con el estudio de los supuestos básicos del MANOVA. Con estas cuestiones ya tratadas, el proceso avanza con la estimación del modelo MANOVA y la valoración del ajuste global del modelo. Cuando se encuentra un modelo MANOVA aceptable, entonces se deben interpretar los resultados con más detalle. El último paso comprende los intentos para validar los resultados y asegurar generalidad en la población. Las Figura 6.3 (pasos 1-3) y 6.5 (pasos 4-6) nos proporcionan un esquema gráfico del proceso, el cual se discute en detalle en las siguientes secciones.

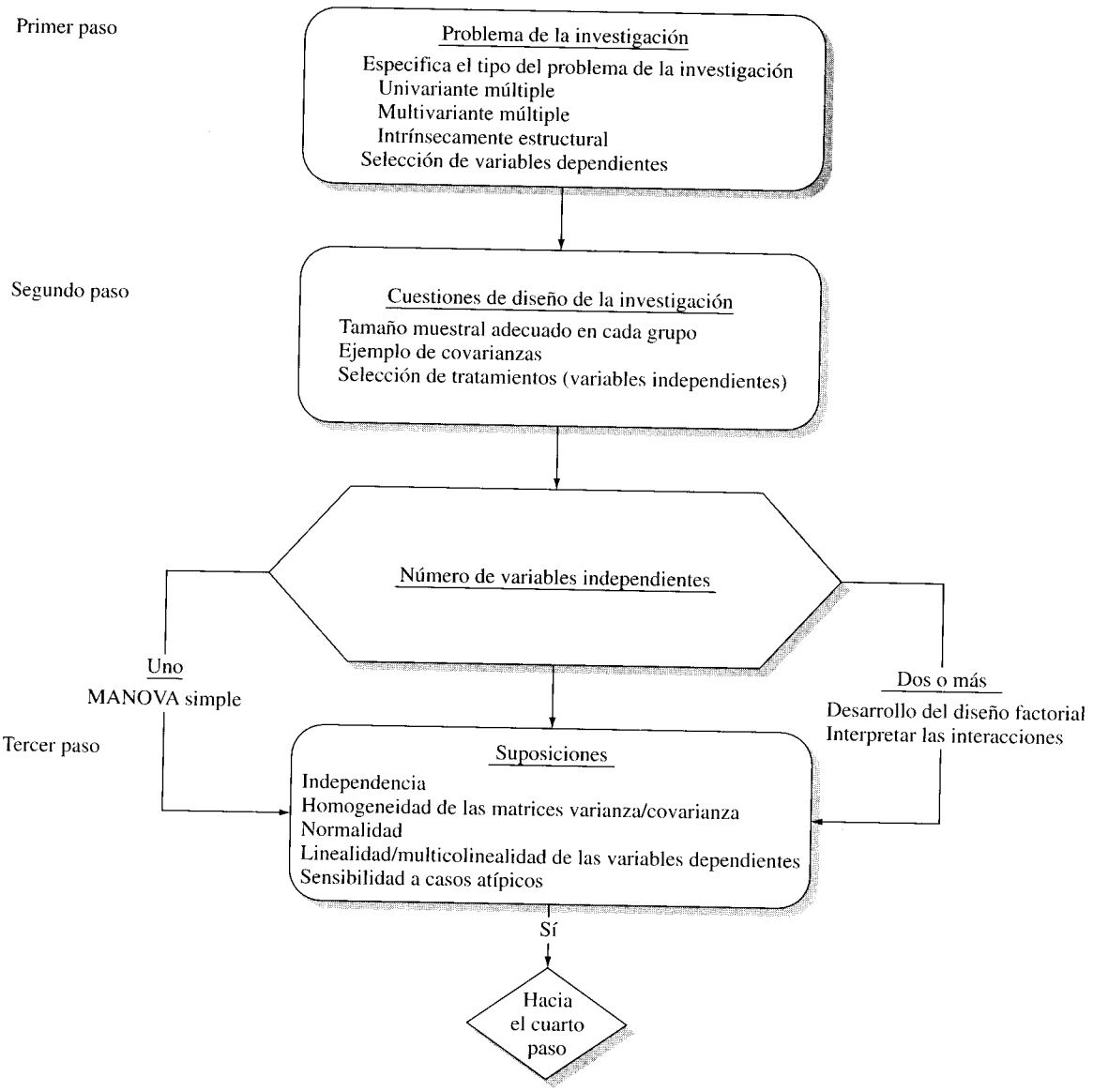


FIGURA 6.3. Pasos 1-3 en el diagrama de decisión del análisis multivariante de la varianza (MANOVA).

Primer paso: Objetivos del MANOVA

La selección del MANOVA está basada en el deseo de analizar una relación de dependencia representada como las diferencias en un conjunto de medidas dependientes a través de una serie de grupos formados por una o más medidas independientes categóricas. Como tal, el MANOVA representa una poderosa herramienta analítica apropiada para una amplia colección de cuestiones de investigación. Si se emplea en situaciones cuasi-experimentales o reales (tales como trabajos de campo o estudios de encuestas donde las medidas independientes son categóricas), puede proporcionar no sólo un entendimiento de la naturaleza y del poder predictivo de las medidas independientes, sino también de la interrelaciones y diferencias observadas en el conjunto de medidas dependientes.

Tipos de preguntas adecuadas en el análisis multivariante del MANOVA

Las ventajas del MANOVA frente a un conjunto de ANOVAs univariantes se dan en el campo estadístico ya discutido anteriormente, y también en su capacidad de proporcionar un único método de contrastar un amplio conjunto de cuestiones multivariantes diferentes. A lo largo del texto, nos centraremos en la naturaleza interdependiente del análisis multivariante. El MANOVA tiene la flexibilidad de permitir que el investigador seleccione los contrastes estadísticos más apropiados para las cuestiones que le afectan. Hand y Taylor [9] han clasificado los problemas multivariantes en tres categorías, cada una de las cuales emplea diferentes aspectos del MANOVA en su resolución. Estas tres categorías son preguntas de análisis univariante múltiple, análisis multivariante estructurado y análisis multivariante intrínseco.

Preguntas del análisis univariante múltiple

Un investigador que analice preguntas del análisis univariante múltiple identifica un número de variables dependientes separadas (por ejemplo, edad, renta, educación de los consumidores) que son analizadas de forma separada pero en donde se necesita algún control sobre el porcentaje de error experimental. En este caso, se emplea el MANOVA para valorar si se encuentra alguna diferencia global entre los grupos, y después se llevan a cabo contrastes univariantes separados para dar respuestas individuales a cada variable dependiente.

Preguntas del análisis multivariante estructurado

Un investigador que se ocupa de cuestiones del análisis multivariante estructurado reune dos o más medidas dependientes que tienen unas determinadas relaciones entre ellas. Una situación habitual en esta categoría son las medidas repetidas, donde las respuestas múltiples son reunidas para cada sujeto, quizá a lo largo del tiempo o en una orientación precontraste-postcontraste para algún estímulo, como por ejemplo un anuncio publicitario. En este caso, el MANOVA proporciona un método estructurado para especificar las comparaciones de las diferencias de los grupos sobre un conjunto de medidas dependientes mientras se mantiene la eficiencia estadística.

Preguntas del análisis multivariante intrínseco

Una pregunta del análisis multivariante intrínseco comprende un conjunto de medidas dependientes en donde la principal cuestión es cómo se diferencian *como un total* entre los grupos. Las diferencias en las medidas dependientes individuales son de menor interés que su efecto conjunto. Un ejemplo es la contrastación de medidas de respuesta múltiple que deben ser consistentes, tales como opinión, preferencia e intención de compra, todas ellas relacionadas con diferentes campañas publicitarias. La potencia completa del MANOVA se utiliza en este caso para valorar no solamente las diferencias globales sino también las diferencias entre las combinaciones de medidas dependientes que de otro modo no serían evidentes. Este tipo de cuestión se aborda de forma correcta con el MANOVA debido a su capacidad para detectar diferencias multivariantes, incluso cuando ningún contraste univariante muestra diferencias.

Selección de las medidas dependientes

Al identificar las preguntas apropiadas para el uso del MANOVA, es también importante discutir brevemente el desarrollo de las cuestiones de investigación, específicamente la selección de las medidas dependientes. Un problema común que se da en el MANOVA es la tendencia de los in-

vestigadores a emplear erróneamente uno de sus argumentos —la capacidad de manejar medidas dependientes múltiples— incluyendo variables sin una base teórica o conceptual. El problema se produce cuando los resultados indican que un subconjunto de variables dependientes tiene la capacidad de influir sobre las diferencias globales entre los grupos. Si algunas de las medidas dependientes con importantes diferencias no son realmente adecuadas para dar respuesta a la cuestión de investigación, entonces diferencias *«falsas»* pueden llevar al investigador a realizar conclusiones incorrectas sobre el *conjunto como un total*. Por ello, siempre se deben examinar a fondo las medidas dependientes y estar seguro de que hay una sólida razón para incluirlas. Debería ser tenida en cuenta cualquier ordenación de las variables, así como los efectos secuenciales posibles. El MANOVA proporciona un contraste especial, el análisis de reducción, para valorar las diferencias estadísticas de una manera secuencial, muy parecido a la inclusión de variables en un análisis de regresión.

En resumen, el investigador debe valorar todos los aspectos de la pregunta de investigación cuidadosamente y asegurarse de que el MANOVA se aplica de la forma más correcta y potente. En las siguientes secciones, nos centramos en muchas cuestiones que tienen un impacto sobre la validez y la precisión del MANOVA; sin embargo, al final es la responsabilidad del investigador la que debe tenerse en cuenta para emplear la técnica correctamente.

Segundo paso: Cuestiones básicas del diseño de la investigación mediante MANOVA

Aunque el MANOVA contrasta los supuestos básicos de la misma forma que el ANOVA y comparte los mismos principios, algunos aspectos de su aplicación son únicos. Estas cuestiones afectan tanto al diseño como a la contrastación estadística del modelo MANOVA.

Requisitos de tamaño muestral — en conjunto y por grupos

Como se podía esperar, el MANOVA requiere tamaños muestrales más grandes que los ANOVAs univariantes y el tamaño muestral debe exceder umbrales específicos en cada celda (grupo) del análisis. Un tamaño de celda mínimo es de 20 observaciones, aunque es posible que se necesiten tamaños de celdas mayores para una potencia estadística aceptable. Como mínimo, el tamaño en cada celda debe ser más grande que el número de variables dependientes incluidas. Aunque este aspecto puede parecer menor, la introducción de un número pequeño de variables dependientes (entre cinco y diez) en el análisis pone una restricción algunas veces molesta en el conjunto de datos. Este problema se pone de manifiesto particularmente en el campo de la experimentación o del estudio de investigación, donde el investigador tiene menor control sobre la muestra obtenida.

Diseños factoriales — dos o más tratamientos

Muchas veces, el investigador desea examinar los efectos de varias variables independientes o tratamientos en vez de usar solamente un tratamiento único en las pruebas de ANOVA o MANOVA. Un análisis con más de dos tratamientos se denomina **diseño factorial**. En general, un diseño con n tratamientos se denomina diseño factorial de n-vías.

Selección de tratamientos

Los diseños factoriales más habitualmente empleados son esas preguntas de investigación que relacionan dos o más variables independientes no-métricas con un conjunto de variables dependientes.

En estos casos, las variables independientes se han especificado en el diseño del experimento o se han incluido en el diseño del experimento de campo en un cuestionario. Pero en algunos casos, los tratamientos son añadidos después de que el análisis ha sido diseñado. Los tratamientos adicionales de uso más común se introducen como **factor en bloques**, que es una característica no-métrica empleada de forma posterior para dividir a los encuestados con el fin de obtener una homogeneidad dentro del grupo más grande y reducir la fuente de diferencias intra grupo. Si se realiza esto, la capacidad de los contrastes estadísticos para identificar las diferencias aumenta. Como ejemplo, supongamos que en nuestro ejemplo de los anuncios publicitarios anterior, se descubriera que los hombres reaccionan de forma diferente que las mujeres frente a los anuncios. Si el género es entonces empleado como un factor en bloques, las diferencias entre los mensajes pueden llegar a ser más evidentes, mientras se tenía que las diferencias estaban ocultas cuando se consideraba que los hombres y las mujeres reaccionaban de forma similar y no estaban separados. Los efectos del tipo de mensaje y el sexo son entonces evaluados de forma separada, proporcionado un contraste más preciso de sus efectos individuales.

Un ejemplo hipotético

Como ejemplo de un diseño factorial simple de dos tratamientos, supongamos que un productor de cereales para el desayuno desea examinar el impacto de tres posibles colores (rojo, azul y verde) y tres formas diferentes (estrellas, cubos y bolas) sobre la evaluación global de un consumidor de un nuevo cereal. Podríamos examinar el impacto de estas variables independientes simultáneamente empleando un diseño factorial 3×3. Los encuestados serían asignados aleatoriamente para evaluar una de las nueve posibles combinaciones de color y forma (usando, por ejemplo, una escala de valoración general de diez puntos). En el análisis de este diseño, se pueden contrastar tres efectos diferentes globales con el ANOVA:

1. El efecto principal del color: ¿existen algunas diferencias entre las clasificaciones medias dadas al rojo (por ejemplo, incluyendo todas las clasificaciones de estrellas rojas, cubos rojos y bolas rojas), con respecto al azul y al verde?
2. El efecto principal de la forma: ¿existen algunas diferencias entre las clasificaciones medias dadas a las estrellas (por ejemplo, incluyendo todas las clasificaciones de estrellas rojas, estrellas azules y estrellas verdes), con respecto a los cubos y a las bolas?
3. El efecto interacción del color y la forma: al igual que la diferencia global entre los colores, ¿es esta diferencia la misma cuando se examine de forma separada de las estrellas, los cubos y las bolas? Por ejemplo, si el rojo fue clasificado muy alto pero recibió una clasificación muy baja cuando fue clasificado como bola (en relación al azul y al verde), este resultado daría evidencia de un efecto interacción; es decir, el efecto del color depende de con qué forma sea considerado. Podríamos plantear esta cuestión de la interacción de una forma equivalente preguntando si el efecto de la forma depende de con qué color es considerado.

En los diseños factoriales del ANOVA, cada uno de estos tres efectos sería contrastado con un estadístico F. Los diseños factoriales del MANOVA son una sencilla extensión del ANOVA; es decir, para cada estadístico F en el ANOVA que evalúa un efecto sobre una única variable dependiente, existe un correspondiente estadístico multivariante (por ejemplo, mrc o lambda de Wilks) que evalúa el mismo efecto sobre un conjunto (vector) de medias de las variables dependientes.

Interpretación de los términos de la interacción

El término interacción representa el efecto conjunto de dos tratamientos y es el efecto que debe ser examinado en primer lugar. Si el efecto interacción no es estadísticamente significativo, entonces los efectos de los tratamientos serán independientes. La independencia en los diseños factoriales

significa que el efecto de un tratamiento es el mismo para cada nivel del(los) otro(s) tratamiento(s), y que los efectos principales pueden ser interpretados directamente. Si la interacción es significativa, entonces se debe determinar el tipo de interacción. Las interacciones pueden ser consideradas como **ordinal** o **disordinal**. Una interacción ordinal se produce cuando los efectos de un tratamiento no son iguales para todos los niveles del otro tratamiento, pero la magnitud es siempre de la misma dirección. En una interacción disordinal, los efectos de un tratamiento son positivos para algunos niveles y negativos para los otros niveles del otro tratamiento.

Las diferencias entre las interacciones son mejor descritas gráficamente. En la Figura 6.4, empleamos el ejemplo de las formas y los colores de los cereales descrito anteriormente. El eje *x* representa las tres categorías de color (rojo, azul y verde). Las líneas conectan la media de la categoría para cada forma a través de los tres colores. Por ejemplo en el gráfico superior, el valor para las bolas rojas es aproximadamente 4,0, el valor para la bola azul es aproximadamente 5,0 y el valor se incrementa ligeramente hasta aproximadamente 5,5 para las bolas verdes.

Se muestran tres formas de interacción. En el caso A, no existe interacción. La falta de interacción se puede ver cuando las líneas que representan las diferencias de las distintas formas entre los niveles del color son paralelas (se observaría el mismo efecto si se representasen las diferencias en el color entre los tres tipos de formas). En el caso donde no existe interacción, los efectos de cada tratamiento son constantes en cada nivel y las líneas son aproximadamente paralelas. En el caso B, vemos que los efectos de cada tratamiento no son constantes y por ello las líneas no son paralelas. Las diferencias para el rojo son grandes, pero disminuyen ligeramente para el cereal azul e incluso más para el cereal verde. Por ello, las diferencias en color varían entre las formas. Pero la ordenación relativa entre los niveles de la forma es la misma, con las estrellas siempre como la más alta, seguidas de los cubos y después de las bolas. Finalmente, en el caso C, las diferencias en el color no sólo varían en magnitud sino también en dirección. Este efecto se observa cuando las líneas no son paralelas y se cortan entre los niveles. Por ejemplo, los cubos tienen una valoración mayor que las estrellas cuando el color es el rojo, pero la valoración es menor para los colores azul y verde.

Si las interacciones significativas son ordinales, el investigador debe interpretar la interacción y asegurar que sus resultados son conceptualmente aceptables. Si es así, entonces se pueden describir los efectos de cada tratamiento. Pero si la interacción significativa es disordinal, entonces los efectos principales de los tratamientos no pueden ser interpretados y el estudio debe ser rediseñado, ya que con la interacción disordinal los efectos varían, no solamente entre los niveles del tratamiento, sino también en la dirección (positiva o negativa). Por ello, los tratamientos no representan un efecto consistente.

Homólogos del MANOVA con otros diseños del ANOVA

Muchos tipos de diseños del ANOVA son tratados en textos de diseños experimentales estándar [13, 16, 18]. Para cada diseño ANOVA, existe un homólogo multivariante; es decir, cualquier ANOVA sobre una variable dependiente se puede extender a diseños del MANOVA. Para ilustrar esto, tendríamos que discutir cada diseño del ANOVA en detalle. Claramente, esto no es posible en un solo capítulo, porque existen libros enteros en donde se tratan los diseños del ANOVA. Para obtener más información, se remite al lector a textos más estadísticamente orientados [1, 2, 5, 6, 7, 8, 10, 17, 21] .

Uso de covariaciones— ANCOVA y MANCOVA

En cualquier diseño del ANOVA se pueden incluir variables independientes métricas, denominadas **covarianzas**. El diseño es entonces calificado como diseño del análisis de la covarianza (ANCOVA). Las covarianzas métricas son incluidas normalmente en un diseño experimental para eliminar influencias extrañas de la variable dependiente, que incrementen la varianza dentro de los grupos (CM_I). Esta situación es similar al uso de un factor de bloqueo, solamente que esta vez la

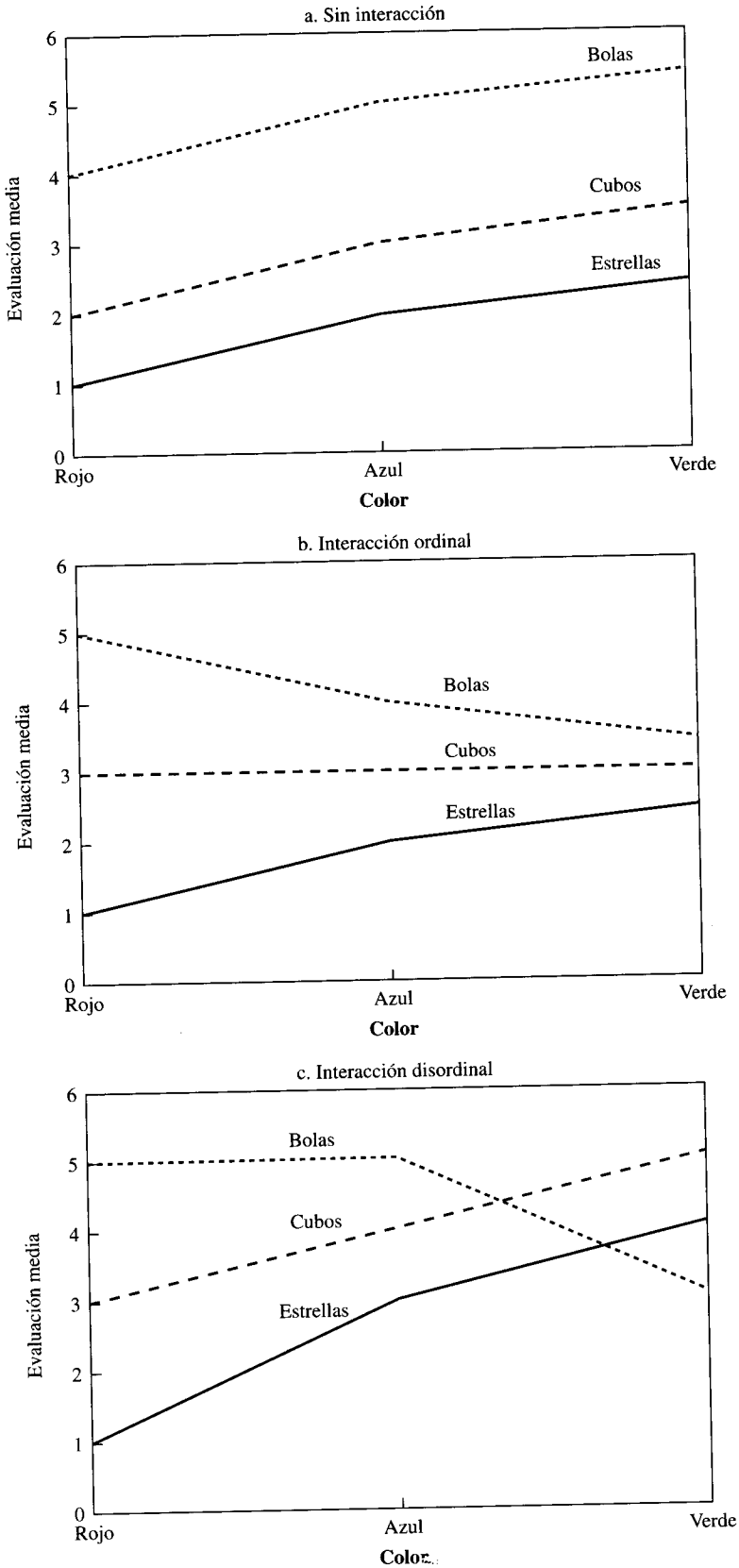


FIGURA 6.4. Efectos de interacción en los diseños factoriales.

variable es métrica. Se han empleado procedimientos similares a la regresión lineal para eliminar la variación de la variable dependiente asociada con una o más covarianzas. Después se lleva a cabo un análisis ANOVA convencional sobre la variable dependiente ajustada. El análisis multivariante de covarianza (MANCOVA) es una extensión simple de los principios de ANCOVA para el análisis multivariante (variables dependientes múltiples); es decir, se puede considerar MANCOVA como MANOVA de los residuos de regresión, o varianza en las variables dependientes que no son justificados por las covarianzas.

Objetivos del análisis de la covarianza

El **análisis de covarianzas** es apropiado para lograr dos objetivos específicos: (1) eliminar cualquier error sistemático fuera del control del investigador que pueda sesgar los resultados, y (2) tener en cuenta las diferencias en las respuestas debidas a características propias de los encuestados. Un sesgo sistemático puede ser eliminado por medio de la asignación aleatoria de los encuestados a varios tratamientos. Sin embargo, en estudios no experimentales, estos controles no son posibles. Por ejemplo, al contrastar los anuncios publicitarios, los efectos pueden diferir dependiendo del momento del día o de la composición de la audiencia y de sus reacciones. El objetivo de la covarianza es eliminar cualquiera de los efectos que (1) influyen solamente a una parte de los encuestados, (2) varían entre los encuestados. Por ejemplo, las diferencias personales, tales como actitud u opiniones, pueden afectar a las respuestas, pero el experimento no las incluye como un factor de tratamiento. El investigador utiliza una covarianza para extraer cualquiera de las diferencias debidas a estos factores antes de que los efectos del experimento sean calculados. Este es el segundo papel del análisis de la covarianza.

Selección de las covarianzas

Una covarianza efectiva en el ANCOVA es aquella que está altamente correlacionada con la variable dependiente pero no está correlacionada con las variables independientes. La varianza de la variable dependiente forma la base de nuestro término de error en el ANOVA. Si nuestra covarianza está correlacionada con la variable dependiente, podemos explicar algo de la varianza (por medio de la regresión lineal), y nos quedamos solamente con la varianza residual de la variable dependiente que no hubiera sido explicado por la variable independiente de todas formas (porque la covarianza no está correlacionada con la variable independiente). Esta varianza residual proporciona un término de error más pequeño (CM_I) para el estadístico F y por ello un contraste más eficiente para los efectos del tratamiento.No obstante, si se correlaciona la covarianza con la variable independiente, entonces la covarianza «explicará» parte de la varianza que podría haber sido «explicada» por la variable independiente y reducir sus efectos. Dado que se extrae primero la covarianza, cualquier variación que se asocia con la covarianza no está disponible para las variables independientes.

Una cuestión habitual es determinar cuántas covarianzas se añaden al análisis. Aunque el investigador quiera tener en cuenta tantos efectos extraños como sea posible, un número demasiado grande reducirá la eficiencia estadística de los procedimientos. A ojo de buen cubero [12] el número de covarianzas debe ser menor que (0,10 × tamaño muestral) – (número de grupos –1). Por ejemplo, para un tamaño muestral de 100 encuestados y 5 grupos, el número de covarianzas debe ser menor que 6, esto es, 0,10 × 100 – (5 – 1). El investigador siempre debe intentar minimizar el número de covarianzas, mientras también asegura las que no van a ser eliminadas, ya que en muchos casos, particularmente cuando hay tamaños muestrales pequeños, estas covarianzas pueden mejorar de manera importante la sensibilidad de los contrastes estadísticos.

Existen dos requisitos para el empleo del análisis de la covarianza: (1) las covarianzas deben tener algún tipo de relación con las medidas dependientes, y (2) las covarianzas deben presentar una homogeneidad del efecto de la regresión, significando esto que la(s) covarianza(s) tiene(n) efectos iguales sobre la variable dependiente entre los grupos. Existen contrastes estadísticos para va-

lorar si este supuesto básico es cierto para cada covarianza empleada. Si alguno de estos requisitos no se da, entonces el empleo de la covarianza es inapropiado.

Un caso especial del MANOVA: Medidas repetidas

Hemos tratado un conjunto de situaciones en las que deseamos examinar las diferencias sobre varias medidas dependientes. Una situación especial de este tipo se produce cuando el mismo encuestado proporciona varias medidas, tales como puntuaciones de un contraste a lo largo del tiempo, y nosotros deseamos examinarlas para comprobar si existe alguna tendencia. Sin embargo, sin un tratamiento especial, estaríamos violando el supuesto básico más importante, la independencia. Existen modelos MANOVA especiales, denominados **medidas repetidas**, que pueden tener en cuenta esta dependencia y sirven para averiguar si existen algunas diferencias entre los individuos del conjunto de variables dependientes. La perspectiva de una persona-de-dentro es importante para que cada persona sea considerada en «igualdad»; por ejemplo, supongamos que estuviéramos valorando la mejora en las puntuaciones de un contraste a lo largo de un semestre. Debemos tener en cuenta las puntuaciones de los contrastes anteriores y cómo éstas se relacionan con las puntuaciones posteriores, y podríamos esperar que se observaran diferentes tendencias para aquellos con puntuaciones iniciales bajas frente a los que las tienen altas. Por ello, debemos «emparejar» las puntuaciones de cada encuestado cuando se lleva a cabo el análisis. No trataremos los detalles de los modelos de medidas repetidas en este texto, ya que es una forma especial del MANOVA. Se remite al lector interesado a un buen número de textos excelentes sobre el tema y a los paquetes estadísticos disponibles [1, 2, 5, 6, 7, 8, 10, 17, 21].

Tercer paso: Supuestos básicos del ANOVA y del MANOVA

Los procedimientos de los contrastes univariantes del ANOVA descritos en este capítulo son válidos (en un sentido formal) solamente si se supone que la variable dependiente está distribuida normalmente y que las varianzas son iguales para todos los grupos de tratamiento. Sin embargo, existe evidencia [16, 23] de que los contrastes F en el ANOVA son robustos respecto a estos supuestos excepto en algunos casos extremos. Para que los procedimientos de los contrastes multivariantes del MANOVA sean válidos, se deben cumplir tres supuestos: (1) las observaciones deben ser independientes, (2) las matrices de varianzas-covarianzas deben ser iguales para todos los grupos de tratamiento y (3) el conjunto de las p-variables dependientes debe seguir una distribución normal multivariante (por ejemplo, cualquier combinación lineal de las variables dependientes debe seguir una distribución normal) [10]. Además de los estrictos supuestos estadísticos, el investigador también debe considerar varios aspectos que afectan a los posibles efectos —a saber, la linealidad y la multicolinealidad del valor teórico de las variables dependientes.

Independencia

El incumplimiento más básico, aunque más importante, de un supuesto ocurre cuando existe una falta de **independencia** entre las observaciones. Existe un número de situaciones tanto experimentales como no experimentales en las cuales este supuesto puede ser violado fácilmente. Por ejemplo, se puede producir un efecto de ordenación en el tiempo (correlación serial) si las medidas son tomadas a lo largo del tiempo, incluso de diferentes encuestados. Otro problema habitual es la obtención de información en la composición de los grupos, por lo que una experiencia común (como una habitación ruidosa o un conjunto confuso de instrucciones) produciría un subconjunto de individuos (aquellos con experiencias comunes) que tendrían respuestas que estarían un tanto

correlacionadas. Finalmente, los efectos extraños y no medidos pueden afectar a los resultados creando una dependencia entre los encuestados. Aunque no existen contrastes con una certeza absoluta para detectar todas las formas de dependencia, el investigador debe explorar todos los posibles efectos y corregirlos si los encuentra. Si se encuentra dependencia entre los grupos de encuestados, entonces una posible solución es combinar a éstos dentro de los grupos y analizar la puntuación media del grupo en lugar de las puntuaciones de los diferentes encuestados. Otro enfoque consiste en emplear alguna forma de análisis de la covariación para tener en cuenta la dependencia. En cualquier caso, o cuando se sospecha que existe dependencia, el investigador debe emplear un menor nivel de significación (0,01 o incluso menor).

Igualdad de las matrices de varianzas-covarianzas

El segundo supuesto básico del MANOVA es la igualdad de las matrices de covarianzas entre los grupos. En este caso, al igual que con el problema de la heteroscedasticidad que tratamos en la regresión múltiple, estamos interesados en las diferencias sustanciales en la cantidad de varianza de un grupo en comparación con la de otro grupo para las mismas variables. Sin embargo, en el MANOVA el interés se centra en las matrices de varianzas-covarianzas de las medidas dependientes para cada grupo. El requisito de igualdad es un contraste muy riguroso, ya que en lugar de contrastar igualdad de varianzas para una única variable como en el ANOVA, el contraste del MANOVA examina todos los elementos de la matriz de covarianzas. Por ejemplo, para cinco variables dependientes, se contrasta la igualdad de las cinco correlaciones y de las diez covarianzas entre los grupos. Afortunadamente, la violación de este supuesto tiene un mínimo impacto si los grupos son aproximadamente de igual tamaño (si el tamaño del grupo más grande dividido por el tamaño del grupo más pequeño es menor de 1,5). Si los tamaños difieren más que esta medida, entonces el investigador, si es posible, debe contrastar y corregir la posible existencia de varianzas distintas. Los programas del MANOVA proporcionan el contraste para la igualdad de las matrices de covarianzas —normalmente el **contraste de Box**— y aportan los niveles de significación para el contraste. Si el investigador encuentra una diferencia significativa que requiere algún tipo de tratamiento, se utiliza una de las muchas transformaciones de estabilización de la varianza disponibles. Se remite al lector al Capítulo 2 para el tratamiento de estos enfoques. Aun así, el contraste de Box es muy sensible a la falta de normalidad [10,19]. Por ello se debe siempre comprobar la normalidad univariante de todas las medidas dependientes antes de llevar a cabo este contraste.

Si persisten las diferencias entre las varianzas después de la transformación y los tamaños de los grupos difieren de forma importante, el investigador debe realizar una serie de ajustes que tengan en cuenta estos efectos. Primero, debemos determinar qué grupo presenta la varianza mayor. Esta determinación se realiza fácilmente, bien examinando la matriz de varianzas-covarianzas o bien usando el determinante de esta matriz, que es calculado por todos los paquetes estadísticos. Si las mayores varianzas pertenecen a los grupos de mayor tamaño, el nivel alfa de significación tiende a ser más grande. Este resultado significa que las diferencias deben ser realmente tenidas en cuenta empleando para ello algún valor más pequeño (por ejemplo, emplear 0,03 en lugar de 0,05). Si las mayores varianzas se dan en los grupos de tamaño más pequeño, entonces se producirá lo contrario. Es decir, la potencia del contraste se vería reducida, y por ello el investigador debería incrementar el nivel de significación.

Normalidad

El último supuesto básico se refiere a la normalidad de las medidas dependientes. En estricto sentido, el supuesto es que todas las variables siguen una **distribución normal multivariante**. La normalidad multivariante supone que el efecto conjunto de dos variables se distribuye normalmente. Aunque este supuesto subyace en la mayoría de las técnicas multivariantes, no existe un contraste preciso para la normalidad multivariante. Por ello, la mayoría de los investigadores contrastan la normalidad univariante para cada variable. Pero aunque la normalidad univariante no garantiza

la normalidad multivariante, si todas las variables cumplen ese requisito, entonces cualquier posible incumplimiento de este supuesto es generalmente insignificante. Las violaciones de este supuesto tienen una pequeña influencia si los tamaños muestrales son grandes, al igual que ocurre en el ANOVA. La violación de este supuesto crea principalmente problemas para utilizar el contraste de Box, pero las transformaciones pueden corregir estos problemas en la mayoría de las situaciones. (Para un debate sobre la transformación de variables, se remite al lector al Capítulo 2). Cuando los tamaños muestrales son medianos, los incumplimientos pequeños pueden ser suavizados siempre y cuando las diferencias sean debidas a asimetria y no a datos anómalos.

Linealidad y multicolinealidad entre las variables dependientes

Aunque el MANOVA valora las diferencias entre combinaciones de medidas dependientes, solamente se puede construir una relación lineal entre las medidas dependientes (y cualquier covariación, si está incluida). Se anima una vez más al investigador, en primer lugar, a examinar los datos, valorando la posible presencia de relaciones no lineales. Si existen, entonces se debe tomar la decisión de si se deben incluir dentro del conjunto de variables dependientes, que tiene como coste aumentar la complejidad pero también se incrementa la representatividad. El Capítulo 2 aborda estas pruebas.

Además del requisito de la linealidad, las variables dependientes no deben presentar alta multicolinealidad entre ellas (se trató en el Capítulo 4), porque esto sólo indica medidas dependientes redundantes y tiende a disminuir la eficiencia estadística. Abordamos el impacto de la multicolinealidad sobre la potencia estadística del MANOVA en la próxima sección.

Sensibilidad a los datos atípicos

Junto con el efecto de la heteroscedasticidad discutido anteriormente, el MANOVA (y el ANOVA) es especialmente sensible a la presencia de datos atípicos, que influyen sobre el error Tipo I. Se incita al investigador, en primer lugar, a examinar los datos en busca de datos atípicos, y a eliminarlos del análisis si es posible, ya que su influencia será muy importante en el conjunto de los resultados.

Cuarto paso: Estimación del modelo MANOVA y valoración del ajuste global

Una vez que el análisis del MANOVA ha sido formulado y los supuestos básicos contrastados, se puede realizar la valoración de las diferencias significativas entre los grupos formados por el(los) tratamiento(s) (véase Figura 6.5). Con esta valoración, el investigador debe seleccionar los contrastes estadísticos más apropiados para los objetivos que se persiguen. Además, en cualquier situación, pero especialmente cuando el análisis llega a ser más complejo, el investigador debe evaluar la potencia de los contrastes estadísticos para proporcionar la mejor representación de los resultados obtenidos.

Criterios para la contrastación de la significación

El MANOVA proporciona varios criterios con los que valorar las diferencias multivariantes entre los grupos. Los cuatro más conocidos son: la mayor raíz característica de Roy, el lambda de Wilks (también conocido como el **estadístico-*U***), la traza de Hotelling y el criterio de Pillai. Como recordará de las discusiones del análisis discriminante en el Capítulo 5, estos criterios valoran las diferencias entre «dimensiones» de las variables dependientes. La mayor raíz característica de Roy

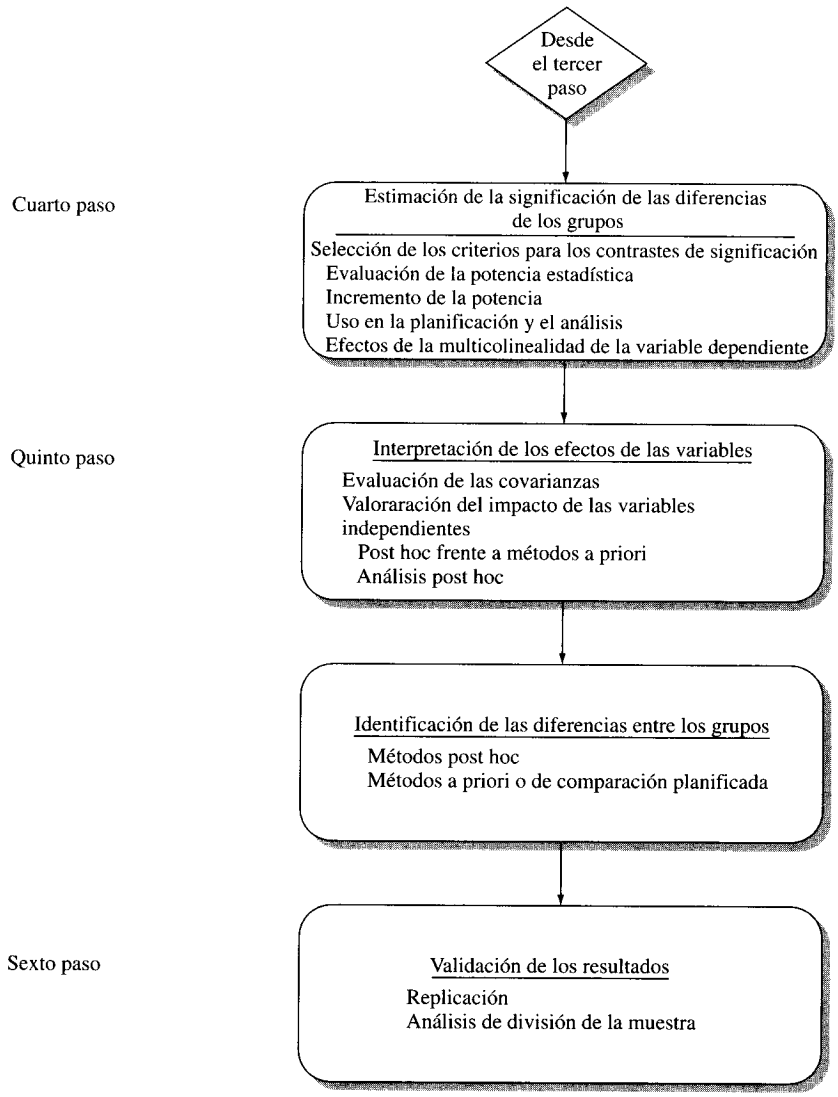


FIGURA 6.5. Pasos 4-6 en el diagrama de decisión del análisis multivariante de la varianza (MANOVA).

como su nombre indica, mide las diferencias solamente sobre la primera raíz canónica (o función discriminante) entre las variables dependientes. Este criterio proporciona alguna ventaja sobre potencia y especificidad de los contrastes, pero los hace menos útiles en ciertas situaciones donde todas las dimensiones deben ser consideradas. El contraste de la mrc de Roy es el más adecuado cuando las variables dependientes están fuertemente interrelacionadas en una sola dimensión, pero también es la medida que es más probable que se vea afectada gravemente por incumplimientos de los supuestos básicos.

Las otras tres medidas valoran todas las posibles fuentes de diferencia entre los grupos. Los lectores que están más familiarizados con el MANOVA por otros textos o por programas estadísticos, probablemente conocen que el contraste más comúnmente empleado para la significación global del MANOVA es el **lambda de Wilks**. Nosotros hemos mencionado la mayor raíz característica y la primera función discriminante, por tanto estos términos implican que pueden existir otras raíces características y otras funciones discriminantes. En realidad, existen p o $(k-1)$ (cualquiera que sea el menor) raíces características o funciones discriminantes, donde p es el número de variables dependientes y k es el número de grupos. A diferencia del estadístico mrc, que está basado en la primera (mayor) raíz característica, el lambda de Wilks considera todas las raíces características; es decir, compara si los

grupos son de algún modo diferentes sin estar afectados por el hecho de que los grupos difieran en al menos una combinación lineal de las variables dependientes. El lambda de Wilks es mucho más fácil de calcular que el estadístico mrc. Su formulación es $|W| / |W + A|$ donde $|W|$ es el determinante (un solo número) de la matriz de dispersión multivariante dentro de los grupos, y |W + A| es el determinante de la suma de W y A donde A es la matriz de dispersión multivariante entre los grupos. Cuanto mayor es la dispersión entre los grupos, más pequeño es el valor del lambda de Wilks y mayor la significación. Aunque la distribución del lambda de Wilks es compleja, se tienen buenas aproximaciones para contrastar la significación, transformándolo en un estadístico F [18].

¿Qué estadístico se prefiere? Los investigadores tienen estos dos estadísticos más un conjunto de posibles medidas que también pueden elegir. Otras medidas ampliamente empleadas son el criterio de Pillai y la traza de Hotelling, ambas similares al lambda de Wilks, dado que consideran todas las raíces características y pueden ser aproximadas por un estadístico F. La medida a utilizar será la más invulnerable a las violaciones de los supuestos básicos del MANOVA, y la que aún mantenga la mayor potencia. Existe un acuerdo entre los expertos en cuanto a que el criterio de Pillai o el lambda de Wilks son los que mejor cumplen estas requisitos, aunque la evidencia sugiere que el criterio de Pillai es más robusto y debe ser empleado si los tamaños muestrales disminuyen, si se dan tamaños de celdas distintos o si se incumple la homogeneidad de las covarianzas. Sin embargo, si el investigador está convencido de que todos los supuestos se cumplen estrictamente, y que las medidas dependientes son representativas de un sola dimensión de los efectos, entonces la mrc de Roy es el estadístico más potente. La mayoría de los paquetes estadísticos incluyen las tres medidas, y se puede realizar una comparación entre ellas.

Potencia estadística de los contrastes multivariantes

En términos sencillos, la **potencia** es la probabilidad de que el contraste estadístico identifique un efecto del tratamiento si este realmente existe. La potencia puede ser definida como uno menos la probabilidad del **error de Tipo II** (β). Como tal, la potencia está relacionada con el nivel de significación alfa (α), que define el error de Tipo I aceptable. Para una discusión más detallada sobre los errores de Tipo I y Tipo II y la potencia, se remite al lector al Capítulo 1.

El nivel de potencia para cualquiera de los cuatro criterios estadísticos —el mrc de Roy, el lambda de Wilks, la traza de Hotelling o el criterio de Pillai— se basa en tres cuestiones: el nivel de significación alfa, el efecto tamaño del tratamiento y el tamaño muestral de los grupos. La potencia está inversamente relacionada con el nivel de significación alfa seleccionado. A medida que alfa se incrementa (es más cauteloso, pasando de 0,05 a 0,01), la potencia disminuye. Por tanto, si el investigador reduce el nivel de significación alfa para reducir el error de Tipo I, como en el caso de una posible dependencia entre las observaciones o un ajuste para comparaciones múltiples, la potencia disminuye. El investigador siempre debe conocer las implicaciones del ajuste del nivel de significación alfa, ya que el objetivo del análisis que no se tiene en cuenta, no es solamente evitar los errores de Tipo I, sino también identificar los efectos del tratamiento si en verdad existen. Si se impone un nivel de significación alfa demasiado riguroso, entonces la potencia puede ser demasiado baja para identificar resultados válidos. El investigador debe considerar, no sólo el nivel de significación alfa, sino también la potencia resultante, y que intente mantener un nivel de significación alfa aceptable con una potencia cerca de 0,80.

El incremento de potencia en ANOVA y MANOVA

Pero, ¿cómo puede el investigador incrementar la potencia una vez que el nivel de significación alfa ha sido especificado? La principal «herramienta» que tiene a su disposición es el tamaño muestral de los grupos. Pero antes de valorar el papel del tamaño muestral, necesitamos comprender el impacto del **efecto tamaño**, que es una medida estandarizada de las diferencias de los grupos, generalmente definida como las diferencias en las medias de los grupos divididas por sus desviaciones estándar. La magnitud del efecto tamaño tiene un efecto directo sobre la potencia del contraste es-

tadístico. Para cualquier tamaño muestral dado, la potencia del contraste estadístico será mayor cuanto mayor sea el efecto tamaño. Contrariamente, si el tratamiento tiene un efecto tamaño esperado más pequeño, se debe conseguir un tamaño muestral mayor para lograr la misma potencia que un tratamiento con un mayor efecto tamaño.

Una vez especificado el nivel de significación alfa e identificado el efecto tamaño, el último elemento que afecta a la potencia es el tamaño muestral. En muchos casos, este es el elemento sobre el que el investigador tiene menos control. Como se discutió anteriormente, incrementos en el tamaño muestral generalmente reducen el error muestral e incrementan la sensibilidad (potencia) del contraste. En trabajos con tamaños de grupos menores a 50 miembros, obtener niveles de potencia adecuados puede ser bastante problemático. El aumento de los tamaños muestrales en cada grupo tiene efectos importantes, hasta que se alcanzan unos tamaños de aproximadamente 150 miembros, y entonces los incrementos en la potencia disminuyen notablemente. También se debe tener cierta precaución cuando los tamaños muestrales son grandes. En muchos contrastes estadísticos, tamaños muestrales grandes reducen el componente del error muestral a un nivel tan pequeño que cualquier diferencia se considera estadísticamente significativa. Cuando los tamaños muestrales llegan a ser grandes y la significación estadística es la indicada, el investigador debe examinar la potencia y los efectos tamaños para asegurarse, no solamente significación estadística, sino también significación práctica.

La utilización de potencia en la planificación y el análisis

La estimación de la potencia debe ser empleada tanto en planificar el análisis como en la valoración de los resultados. En la etapa de planificación, el investigador determina el tamaño muestral necesario para identificar el efecto tamaño estimado. En muchos casos, el efecto tamaño puede ser estimado a partir de una investigación previa a partir de juicios razonados o incluso ponerlo a un nivel mínimo de significación práctica. En cada caso, se puede determinar el tamaño muestral necesario para lograr un determinado nivel de potencia con un nivel de significación específico.

Al valorar la potencia de los contrastes después de que el análisis ha sido realizado, se está proporcionando un contexto en el que interpretar los resultados, especialmente si no se han encontrado diferencias significativas. El investigador debe determinar primero si la potencia obtenida fue suficiente (0,80 o más). Si no fuese así, ¿el análisis podría ser reformulado para lograr más potencia? Una posibilidad sería realizar alguna forma de tratamiento en bloques o de análisis de la covarianza que hará que el contraste sea más eficiente haciendo énfasis en el efecto tamaño. Si la potencia fuese la adecuada y no se encontrase significación estadística para el efecto del tratamiento, entonces lo más probable es que el efecto tamaño para el tratamiento fuese demasiado pequeño para considerar significación estadística o práctica.

El cálculo de los niveles de potencia

Para calcular la potencia en los análisis ANOVA, se encuentran disponibles tanto fuentes publicadas [3, 20] como programas de computador. Sin embargo, los métodos de calculo de la potencia en los MANOVA están mucho más limitados. Afortunadamente, la mayoría de los programas informáticos proporcionan una valoración de la potencia para los contrastes de significación y permiten al investigador determinar si la potencia debe jugar un papel en la interpretación de los resultados. Con respecto al material publicado con objetivos de planificación, existe poco para el MANOVA dado que muchos elementos afectan a la potencia de un análisis MANOVA. Una fuente [15] con tablas publicadas refleja la potencia de un conjunto de situaciones habituales donde se ha aplicado el MANOVA. La Tabla 6.2 proporciona un panorama de los tamaños muestrales necesarios para varios niveles de complejidad del análisis. Un repaso de la tabla nos lleva a destacar varios puntos generales. Primero, un incremento del número de variables dependientes requiere un incremento del tamaño muestral para mantener un nivel de potencia dado. El tamaño muestral adicional que se necesitaría es más pronunciado para los efectos tamaño más pequeños. En segundo lugar, si se

TABLA 6.2. Requisitos de tamaño muestral para obtener una potencia estadística de 0,80 en el MANOVA

	Número de grupos											
	3				4				5			
	Número de variables dependientes				*Número de variables dependientes*				*Número de variables dependientes*			
Efecto tamaño	2	4	6	8	2	4	6	8	2	4	6	8
Muy grande	13	16	18	21	14	18	21	23	16	21	24	27
Grande	26	33	38	42	29	37	44	48	34	44	52	58
Medio	44	56	66	72	50	64	74	84	60	76	90	100
Pequeño	98	125	145	160	115	145	165	185	135	170	200	230

Fuente: Läuter, J. (1978), «Sample Size Requirements for the T^2 Test of MANOVA (Tables for One-Way Classification)», *Biometrical Journal* 20: 389–406.

espera que los efectos tamaño sean pequeños, el investigador debe estar preparado para llevar a cabo un esfuerzo de investigación sustancial para lograr niveles aceptables de potencia. Por ejemplo, para lograr una potencia propuesta de 0,80 cuando se valoran los efectos tamaño en un estudio con cuatro grupos, se requieren 115 individuos si se emplean dos medidas dependientes. El tamaño muestral requerido aumenta a 185 por grupo si se consideran ocho variables dependientes. Los beneficios de la parsimonia en el conjunto de variables dependientes aparecen no solamente en la interpretación sino también en los contrastes estadísticos para las diferencias de los grupos.

Los efectos de la multicolinealidad de la variable dependiente sobre la potencia

Hasta ahora hemos tratado la potencia desde una perspectiva aplicable tanto al ANOVA como MANOVA. Sin embargo, en MANOVA, el investigador también tiene que considerar los efectos de multicolinealidad de las variables dependientes sobre la potencia de los contrastes estadísticos. El investigador, bien en la etapa de planificación o bien en la etapa de análisis, tiene que considerar la fuerza y dirección de las correlaciones además de los efectos tamaño de las variables dependientes. Si clasificamos las variables por sus tamaños de efecto como fuertes o débiles, entonces surgen pautas [4]. En primer lugar, si la pareja de variables correlacionadas está compuesta de variables fuertes-fuertes o débiles-débiles, entonces se obtiene la máxima potencia cuando la correlación entre las variables es altamente negativa. Esto sugiere que el MANOVA está en las mejores condiciones con la adición de las variables dependientes que tienen correlaciones altamente negativas. Por ejemplo, en vez de incluir dos medidas de satisfacción redundantes, el investigador podría reemplazarlas con las medidas de correlación de satisfacción y disatisfacción para incrementar la potencia. Cuando la pareja de variables correlacionada es una mezcla (fuerte-débil), entonces la potencia está maximizada cuando la correlación es alta, bien positiva o negativa. Una excepción es el resultado de que el uso de los ítems múltiples para incrementar la fiabilidad tiene como resultado una ganancia neta de potencia, incluso si los ítems son redundantes y positivamente correlacionados.

Quinto paso: Interpretación de los resultados del MANOVA

Una vez que se ha evaluado la significación estadística de los tratamientos, el investigador puede desear examinar los resultados por medio de cualquiera de estos tres métodos: (1) interpretar los

efectos de las covarianzas si éstas han sido empleadas, (2) valorar qué variable(s) dependiente(s) presenta(n) diferencias entre los grupos, o (3) identificar qué grupos difieren en una sola variable dependiente o en el valor teórico dependiente completo. Primero examinaremos los métodos con los que se identifican las covarianzas significativas y las variables dependientes y después se tratarán los métodos con los que se pueden medir las diferencias entre los diferentes grupos.

Evaluación de las covarianzas

Habiéndose dado los supuestos para la aplicación de las covarianzas, el investigador puede querer interpretar su efecto real sobre el valor teórico dependiente y su impacto sobre los verdaderos contrastes estadísticos de los tratamientos. Dado que el ANCOVA y MANCOVA son aplicaciones de los procedimientos de regresión dentro del análisis del método de la varianza, la valoración del impacto de las covarianzas sobre las variables dependientes es bastante similar a examinar las ecuaciones de regresión. Para cada covarianza se forma una ecuación de regresión que señala la validez de la relación predictiva. Si las covarianzas representan teóricamente efectos despreciables, entonces estos resultados proporcionan una base objetiva para aceptar o rechazar las relaciones propuestas. En la práctica, el investigador puede examinar el impacto de las covarianzas y eliminar aquéllas con poco o ningún efecto.

También se debe examinar el impacto global de la introducción de la(s) covarianza(s) en un contraste estadístico de los tratamientos. El enfoque más directo es realizar el análisis con y sin las covarianzas. Aquellas que sean efectivas mejorarán la potencia estadística de los contrastes y reducirán la varianza dentro del grupo. Si no se observa ninguna mejora sustancial, entonces pueden ser eliminadas, ya que reducen los grados de libertad disponibles para los contrastes de los efectos del tratamiento. Este enfoque también puede servir para identificar aquellos casos en los que la covarianza es «demasiado potente» y reduce la varianza de tal manera que los tratamientos no son significativos; hecho que se produce muchas veces cuando se incluye una covarianza que está correlacionada con una de las variables independientes y de esta manera «elimina» esta varianza, reduciendo por ello la capacidad explicativa de la variable independiente.

Evaluación del valor teórico dependiente

El siguiente paso es el análisis del valor teórico dependiente para evaluar cuál de las variables dependientes contribuye a las diferencias globales señaladas en los contrastes estadísticos. Este paso es esencial ya que se puede identificar un conjunto de variables que, o bien acentúa las diferencias mientras otras variables no son significativas, o bien oculta los efectos significativos de las restantes. Los procedimientos descritos en esta sección se denominan **contrastes post hoc** —constrastes que se realizan después de examinar el modelo de los datos. Otro enfoque es el empleo de **contrastes a priori**— contrastes que se planean con anterioridad a observar los datos desde un punto de vista de toma de decisión teórico o práctico. Desde un punto de vista pragmático, las situaciones surgen cuando una variable dependiente clave debe ser aislada y contrastada con máxima potencia. Recomendamos que se lleve a cabo un contraste a priori en tales situaciones. El enfoque más común consiste en realizar los contrastes univariantes para las variables seleccionadas. Por ejemplo, en el caso de dos grupos, un contraste t ordinario es un contraste a priori para una variable dependiente dada. Sin embargo, los investigadores deben saber que a medida que se incrementa el número de estos contrastes a priori, se invalida una de las principales ventajas del enfoque multivariante para contrastar la significación —el control del porcentaje del error de Tipo I—, a menos que se realicen ajustes específicos que controlen el aumento del error de Tipo I.

Un análisis MANOVA de dos grupos implica un ajuste del estadístico T^2. Cuando el estadístico T^2 excede a T^2_{crit} para un nivel α dado, concluimos que los vectores de las puntuaciones medias son diferentes. La función discriminante nos informa de qué combinación lineal de las variables dependientes produce la diferencia entre los grupos más fiable, pero otras comparaciones también pueden ser de interés. Si quisiéramos contrastar las diferencias de los grupos individualmente para cada

una de las variables dependientes, calcularíamos un estadístico t estándar y lo compararíamos con la raíz cuadrada de T^2_{crit} (por ejemplo, T_{crit}) para juzgar su significación. Este procedimiento aseguraría que la probabilidad de cualquier error de Tipo I entre todos los contrastes debería mantenerse en α (donde α se especificó en el cálculo de T^2_{crit}) [10]. Podríamos realizar contrastes similares en situaciones con k grupos ajustando el nivel α con la **desigualdad de Bonferroni**, la cual establece que el nivel alfa debe ser ajustado según el número de contrastes que se están realizando. El nivel alfa ajustado empleado en cualquier contraste separado se define como el nivel alfa total dividido por el número de contrastes (alfa ajustado = alfa total / número de contrastes).

También se puede emplear un procedimiento conocido como **análisis de reducción** [14,19] para valorar individualmente las diferencias entre las variables dependientes. Este procedimiento comprende el cálculo de un estadístico univariante F para una variable dependiente después de eliminar los efectos de otras variables dependientes que la preceden en el análisis. El procedimiento es de alguna forma similar a la regresión por etapas, pero aquí se examina si una determinada variable dependiente proporciona información única (incorrelacionada) sobre las diferencias de los grupos. Los resultados del análisis de reducción serían exactamente los mismos que los que se obtendrían si se realizase un análisis de covarianzas con las otras variables dependientes precedentes empleadas como covarianzas. Un supuesto decisivo del análisis de reducción es que el investigador conoce el orden en que se deben introducir las variables dependientes, ya que la interpretación puede variar fuertemente según diferentes órdenes para entrar. Si la ordenación tiene un apoyo teórico, entonces el contraste del análisis de reducción es válido. Las variables identificadas como no significativas son «redundantes» con las anteriores variables significativas, con lo que añaden más información referente a las diferencias entre los grupos. Se puede cambiar el orden de las variables dependientes para comprobar si los efectos de las variables son redundantes o únicos, pero el proceso es bastante complicado conforme se incrementa el número de variables.

Otros procedimientos incluyen más análisis de las funciones discriminantes, en particular sobre la primera función discriminante, para obtener información adicional sobre qué variables diferencian mejor entre los grupos. Todos estos análisis están pensados para hacer comprender al investigador cuál de las variables dependientes contribuye a las diferencias en el valor teórico dependiente entre el(los) tratamiento(s).

Identificación de las diferencias entre los distintos grupos

Aunque los contrastes univariantes y multivariantes del ANOVA y del MANOVA nos permiten probar la hipótesis nula de que las medias de los grupos son todas iguales, no nos señalan dónde se establecen esas diferencias significativas. Los contrastes t múltiples no son apropiados para contrastar la significación de las diferencias entre las medias de pares de grupos dado que la probabilidad del error de Tipo I se incrementa con el número de comparaciones realizadas entre grupos (igual que el problema al emplear múltiples ANOVAs univariantes frente al MANOVA). Se encuentran disponibles muchos procedimientos para investigaciones posteriores de diferencias de las medias de grupos específicos de interés, que pueden ser clasificados como a priori o post hoc. Estos procedimientos emplean enfoques diferentes para controlar el porcentaje del error de Tipo I entre los múltiples contrastes.

Métodos post hoc

Entre los métodos post hoc más habituales consideramos (1) contraste de Scheffe, (2) método de la diferencia en verdad significativa (DVS) de Tukey, (3) extensión del enfoque de la diferencia menos significativa (DMS) de Tukey, (4) contraste de rango múltiple de Duncan, y (5) el contraste de Newman-Kuels. Cada método identifica qué comparaciones entre los grupos (por ejemplo, grupo 1 frente a grupo 2 y 3) presentan diferencias significativas. Proporcionan al investigador los contrastes para cada combinación de grupos, simplificando con ello el proceso interpretativo.

Aunque estos métodos simplifican la identificación de las diferencias de los grupos, todos ellos comparten el problema de tener niveles bastante bajos de potencia. Dado que los contrastes post hoc deben examinar todas las posibles combinaciones, la potencia de cualquier contraste individual es bastante baja. A estos cinco contrastes de significación post hoc o de comparación múltiple se les ha contrastado su potencia [19]. Las conclusiones son que el contraste de Scheffe es el más prudente con respecto al error de Tipo I. Los restantes contrastes son clasificados en este orden: DVS de Tukey, DMS de Tukey, Newman-Kuels y Duncan. Si los efectos tamaño son grandes o el número de grupos es pequeño, los métodos post hoc pueden identificar las diferencias de los grupos. Pero el investigador debe tener en cuenta las limitaciones de estos métodos y emplear otros métodos si se pueden identificar comparaciones más específicas. Una discusión de las posibles opciones de cada método está por encima del alcance este capítulo. Se pueden encontrar explicaciones y tratamientos excelentes de estos procedimientos en varios textos [12, 23].

Contrastes a priori o comparaciones planificadas

El investigador también puede realizar comparaciones específicas entre los grupos empleando un contraste a priori o **comparaciones planificadas**. Este método es similar a los contrastes post hoc descritos anteriormente pero se diferencia en que el investigador establece qué comparaciones de los grupos se van a realizar en las comparaciones planificadas frente a contrastar el conjunto completo, como se hacía en los contrastes post hoc. Las comparaciones planificadas son más potentes porque el número de comparaciones es menor, pero tener más potencia resulta de poca utilidad si el investigador no contrasta específicamente las comparaciones de grupos «correctas». Las comparaciones planificadas son más apropiadas cuando los fundamentos conceptuales pueden apoyar las comparaciones específicas que se van a realizar. Las comparaciones planificadas no deben emplearse de una manera explicativa porque no tienen controles efectivos frente al aumento de los niveles del error de Tipo I total.

El investigador establece los grupos que se van a comparar a través de un **contraste**, que es justamente una combinación de las medias de los grupos que representa una comparación planificada específica. Los contrastes pueden ser construidos generalmente como:

$$C = W_1G_1 + W_2G_2 + \ldots + W_kG_k$$

donde

C = valor del contraste
W = ponderaciones
G = medias de los grupos

El contraste se realiza asignando ponderaciones positivas y negativas para especificar los grupos que se van a comparar, mientras se asegura que la suma de las ponderaciones es cero. Por ejemplo, suponga que tenemos tres medias de grupos. Para contrastar la diferencia entre G_1 y G_2, $C = (1)G_1 + (-1)G_2 + (0)G_3$. Para contrastar si la media de G_1 y G_2 es diferente de la de G_3, el contraste se construye como $C = (0{,}5)G_1 + (0{,}5)G_2 + (-1)G_3$. Se calcula un estadístico F separado para cada grupo. De esta manera, el investigador puede definir cualquier comparación deseada y contrastarla directamente, pero la probabilidad del error de Tipo I para cada comparación a priori es igual a α. De esta manera, varias comparaciones planificadas incrementarán el nivel total del error de Tipo I. Todos los paquetes estadísticos pueden realizar tanto contrastes a priori como post hoc para variables dependientes únicas.

Si el investigador desea llevar a cabo comparaciones del valor teórico dependiente completo, están disponibles extensiones de estos métodos. Después de concluir que los vectores de las medias de los grupos no son iguales, el investigador podría estar interesado en ver si existen algunas diferencias de los grupos sobre el valor teórico dependiente compuesto. Se puede calcular un estadístico F ANOVA estándar y compararlo con $F_{crit} = (N - k)\, mrc_{crit} / (k - 1)$, donde el valor de mrc_{crit} está tomado de la distribución de mrc con grados de libertad adecuados.

Sexto paso: Validación de los resultados

Las técnicas del análisis de la varianza (ANOVA y MANOVA) se han desarrollado según la tradición de la experimentación, donde la **replicación** es el principal medio de validación. La especificidad de los tratamientos experimentales permite un amplio empleo del mismo experimento en múltiples poblaciones para evaluar la generalidad de los resultados. Este es el principal dogma del método científico. Sin embargo, en las ciencias sociales y en la investigación económica, la experimentación verdadera es muchas veces sustituida por contrastes estadísticos en situaciones no experimentales como un estudio con encuesta. La capacidad para validar los resultados en estas situaciones se basa en la replicación de los tratamientos. En muchos casos, se usan como tratamientos características demográficas tales como edad, sexo, renta y los gustos. Estos tratamientos puede parecer que cumplen el requisito de comparabilidad, pero el investigador debe asegurarse que el elemento adicional de la asignación aleatoria a las celdas también se cumple; sin embargo, muchas veces en un estudio con encuesta esto no es cierto. Por ejemplo, el uso de la edad y el género como las variables independientes es un ejemplo común del empleo del ANOVA o MANOVA en un estudio con encuesta. Pero en términos de validación, el investigador debe ser precavido al analizar múltiples poblaciones y comparar los resultados como una mera prueba de validez. Dado que los encuestados en un sentido simple «se seleccionan ellos mismos», los tratamientos en este caso no pueden ser asignados por el investigador y con ello la asignación aleatoria no se puede realizar. Por tanto, el investigador debe considerar de manera importante el empleo de covarianzas para controlar otras características que podrían estar relacinadas con los grupos edad/género y que podrían afectar a las variables dependientes, pero que no están incluidas en el análisis.

Otra cuestión es la exigencia de causalidad cuando se emplean métodos o técnicas experimentales. Los principios de la causalidad son examinados con más detalle en el Capítulo 11. Para nuestros objetivos en este capítulo, el investigador debe recordar que en todos los campos de investigación, incluyendo el experimental, ciertos criterios conceptuales (por ejemplo, ordenación temporal de los efectos y resultados) deben ser establecidos antes que la causalidad pueda ser respaldada. La aplicación simple de una técnica en particular en un ámbito experimental no asegura la causalidad.

Resumen

Hemos tratado las aplicaciones adecuadas y las consideraciones importantes del MANOVA para llevar a cabo análisis multivariantes con múltiples medidas dependientes. Aunque existen beneficios considerables en su utilización, el MANOVA debe ser aplicado cuidadosamente y con propiedad en la cuestión que se analice. Cuando así se realiza, el investigador tiene a su disposición una técnica con flexibilidad y potencia estadística. Ahora ilustraremos la aplicación del MANOVA (y su contrapartida univariante, el ANOVA) en una serie de ejemplos.

Ejemplo 1: Diferencia entre dos grupos independientes

Para presentar los beneficios prácticos del análisis multivariante de las diferencias de los grupos, empezamos nuestra discusión con uno de los diseños experimentales más conocidos: el diseño aleatorizado de dos grupos, en donde cada encuestado es aleatoriamente asignado a uno de los dos niveles (grupos) exclusivamente del tratamiento (variable independiente). En el caso univariante, se mide una única variable dependiente métrica, y la hipótesis nula es que los dos grupos tienen medias iguales. En el caso multivariante, se miden múltiples variables dependientes métricas, y la hipótesis nula

es que los dos grupos tienen vectores de medias iguales. En un análisis univariante de dos grupos, el contraste estadístico adecuado es el estadístico *t* (un caso especial del ANOVA); para el análisis multivariante, el contraste estadístico adecuado es la T^2 de Hotelling. Dado que ambos análisis emplean el mismo contexto, no difieren en los enfoques apropiados para la validación, la cual implica esencialmente la replicación en otras muestras o a través de una división de la muestra. Por ello, las cuestiones de validación (sexto paso) no se realizarán para cada uno de los análisis por separado.

Un enfoque univariante: El contraste *t*

Primer paso: *Objetivos del análisis*

De la encuesta HATCO de 100 clientes, a los gerentes les interesaba determinar si HATCO actuaba mejor cuando los clientes utilizaban la especificación global de compra o la especificación por separado (X_{11}) (veáse el Capítulo 1 para una descripción completa de las variables y del estudio completo). Como una medida de la actuación, HATCO decidió usar X_9, el porcentaje del negocio de los clientes que se realiza con HATCO. Cuanto más alto sea el valor, más suministra HATCO sobre un porcentaje base a esa empresa. La hipótesis nula que HATCO desea contrastar es que ellos sirven a ambos tipos de empresas de empresas de igual forma; es decir, el tipo de método de compra no tiene efecto en el nivel de uso (H_0: $\mu_1 = \mu_2$). La hipótesis alternativa es que el método de compra define grupos que compran en proporciones significativamente estadísticas de HATCO (por ejemplo, H_A: $\mu_1 < \mu_2$ o H_A: $\mu_1 > \mu_2$).

Segundo paso: *Diseño de la investigación del ANOVA*

La cuestión principal en el diseño del ANOVA de dos grupos es el tamaño muestral en cada celda. Al igual que en la mayoría de los estudios de encuestas, los tamaños de las celdas son distintos. Al finalizar la encuesta, 40 empresas indicaron que usaban la especificación por separado y 60 empresas usaban el análisis del valor total. Los tamaños de celdas distintos hacen que los contrastes estadísticos sean más sensibles a las violaciones de los supuestos, especialmente el contraste de homogeneidad de la varianza de la variable dependiente. Los investigadores de HATCO no identificaron ninguna variable apropiada para incluirla como covarianza. Finalmente, se consideró inadecuado incluir variables independientes adicionales que creasen un diseño factorial.

Tercer paso: *Supuestos básicos del ANOVA*

La independencia de los encuestados se garantizó tanto como fue posible por medio de un plan de muestreo aleatorio. El supuesto de normalidad y la presencia de datos atípicos para la variable dependiente, X_9, fueron examinados en el Capítulo 2 y se encontró que eran aceptables. Un supuesto particularmente importante para el ANOVA es la homogeneidad de la varianza de la variable dependiente entre los grupos. Existen varios contrastes para contrastar este supuesto (el contraste de Levene, la C de Cochran y el Box de Bartlett). El contraste de Levene indica que no existen diferencias (significación = 0,2434), al igual que el de la C de Cochran (significación = 0,396) y el contraste de Box-Barlett (significación = 0,411). Con esto, tamaños de celdas distintos no deben influir en la sensibilidad de los contrastes estadísticos de las diferencias de los grupos.

Cuarto paso: *Estimación del modelo ANOVA y valoración del ajuste global*

El uso, medido en términos de porcentaje, se muestra en la Tabla 6.3 para todas las empresas. Los dibujos de caja de la Figura 6.6 representan los niveles de uso de los encuestados en dos grupos —los clientes que usan la especificación por separado y los clientes que usan el análisis del valor total. Como se puede observar, los clientes que usan la especificación realizan por separado una

TABLA 6.3. Nivel de uso de las empresas utilizando el método de compra (datos básicos para el contraste *t* univariante o un ANOVA de 2 grupos)

Método de compra					
Grupo 1: Compra de especificación		*Grupo 2: Análisis del valor total*			
Empresa	X_9 *(Nivel de uso)*	*Empresa*	X_9 *(Nivel de usol)*	*Empresa*	X_9 *(Nivel de uso)*
2	43,0	1	32,0	51	41,0
3	48,0	5	58,0	55	39,0
4	32,0	7	46,0	56	47,0
6	45,0	9	63,0	58	65,0
8	44,0	11	32,0	59	46,0
10	54,0	12	47,0	61	54,0
13	39,0	14	38,0	62	60,0
24	36,0	15	54,0	63	47,0
27	36,0	16	49,0	64	36,0
30	46,0	17	38,0	66	45,0
31	43,0	18	40,0	67	59,0
34	47,0	19	54,0	69	58,0
35	35,0	20	55,0	72	55,0
36	39,0	21	41,0	73	51,0
37	44,0	22	35,0	74	60,0
39	29,0	23	55,0	76	49,0
40	28,0	25	49,0	77	42,0
41	40,0	26	49,0	78	47,0
45	38,0	28	54,0	80	56,0
48	43,0	29	49,0	81	59,0
52	53,0	32	53,0	82	47,0
53	50,0	33	60,0	84	37,0
54	32,0	38	46,0	88	36,0
57	62,0	42	58,0	90	60,0
60	50,0	43	53,0	91	49,0
65	40,0	44	48,0	92	39,0
68	46,0	46	54,0	93	43,0
70	49,0	47	55,0	95	31,0
71	50,0	49	57,0	97	60,0
75	41,0	50	53,0	100	33,0
79	39,0				
83	41,0				
85	53,0				
86	43,0				
87	51,0				
89	34,0				
94	36,0				
96	25,0				
98	38,0				
99	42,0				
Media	42,100		48,767		
Varianza	60,653		77,405		
Tamaño muestral	40		60		

Cálculo del estadístico *t*

Error estándar*: $\sqrt{\frac{77{,}405}{60} + \frac{60{,}653}{40}} = 1{,}675$

Estadístico *t*: $\frac{48{,}767 - 42{,}100}{1{,}675} = 3{,}980$

* Esta fórmula del error estándar es apropiada para celdas de tamaños iguales o para casos donde las celdas tienen más de 30 observaciones. Para situaciones en las que las celdas tienen tamaños desiguales o pequeños (menos de 30), veáse Stevens [20].

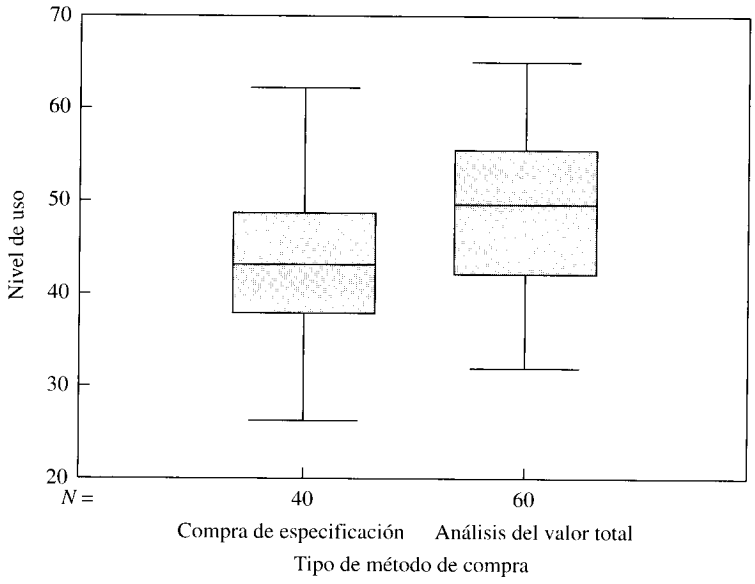


FIGURA 6.6. Dibujos de caja del nivel de uso para un ANOVA de dos grupos.

media del 42,1 por ciento de sus negocios con HATCO, mientras que aquellos que usan el análisis del valor total realizan con HATCO el 48,77 por ciento. El análisis del contraste *t* examinará la diferencia entre los grupos y contrastará estadísticamente la igualdad de las medias de los dos grupos.

Para realizar el contraste, primero elegimos un nivel de significación de 0,05 (el máximo porcentaje permitido del error de Tipo I). De esta manera, antes de llevar a cabo el estudio, ya sabemos que 5 de cada 100 veces podríamos concluir que la metodología de la compra tendría un efecto sobre el porcentaje de uso de la empresa cuando realmente no lo tenía. Todos los programas estadísticos calculan automáticamente los niveles de significación de las diferencias. Para determinar el valor de la t_{crit} atenderemos a la distribución de la *t* con 40 + 60 – 2 = 98 grados de libertad y $\alpha = 0{,}05$. Encontramos que $t_{crit} = 1{,}66$. Lo siguiente es calcular el valor de nuestro estadístico *t*. Como se muestra al final de la Tabla 4.1, $t = 3.98$. Dado que este valor excede al t_{crit}, concluimos que el método de compra afecta al porcentaje de uso de la empresa de los productos de HATCO. La potencia estadística del contraste es de 0,97, asegurando que la diferencia encontrada es estadísticamente significativa.

Quinto paso: *Interpretación de los resultados*

La única variable dependiente y la presencia solamente de dos grupos elimina la necesidad de examinar tanto el valor teórico dependiente como las diferencias entre los grupos, además del conjunto de contrastes descritos en el cuarto paso. El investigador puede informar de los porcentajes estadísticamente significativos de los negocios de los clientes de HATCO que utilizan los dos métodos de compra. Las empresas que utilizan el análisis del valor total compran un porcentaje significativamente mayor de sus productos a HATCO que las empresas que utilizan la especificación por separado. Sin embargo, el investigador debe valorar si la diferencia de aproximadamente el 6 por ciento tiene una significación práctica en la toma de decisiones de la dirección.

Un enfoque multivariante: La T^2 de Hotelling

Probablemente sea incierto suponer que se manifestará una diferencia entre cualesquiera dos grupos experimentales solamente en una única variable dependiente. Por ejemplo, dos mensajes publicitarios no solamente pueden producir diferentes niveles de intención de compra sino que también pueden afectar a un conjunto de otros aspectos (potencialmente correlacionados) de la res-

puesta frente a los anuncios (por ejemplo, evaluación global del producto, credibilidad del mensaje, interés, atención). Muchos investigadores tratan esta situación de múltiple criterio por medio de la aplicación repetida de contrastes t univariantes individuales hasta que sean analizadas todas las variables dependientes. Este enfoque tiene importantes deficiencias. Como se discutió anteriormente, considere lo que podría ocurrir al porcentaje del error de Tipo I (incremento cuando se utilizan múltiples contrastes t) y la incapacidad de contrastes t en parejas para detectar diferencias entre combinaciones de las variables dependientes que no son evidentes con los contrastes univariantes.

Primer paso: *Objetivos del MANOVA*

En nuestro ejemplo univariante, HATCO comparó el nivel de uso de las empresas (X_9 es la variable dependiente) utilizando diferentes metodologías de compra (X_{11} es la variable independiente). Para convertir este ejemplo en uno multivariante, necesitamos al menos dos variables dependientes. Supongamos que HATCO estuviese interesado en los niveles de satisfacción de las empresas con los dos enfoques de compra. Ahora seleccionaríamos X_{10} (satisfacción con HATCO) como la segunda variable dependiente (véase Tabla 6.4). La hipótesis nula que HATCO está contrastando ahora es que los vectores de las puntuaciones medias para cada grupo son iguales (por ejemplo, que el método de compra no tiene ningún efecto ni sobre el uso ni sobre la satisfacción).

Segundo paso: *Diseño de la investigación del MANOVA*

La principal cuestión en un MANOVA de dos grupos sigue siendo el tamaño muestral de cada uno de los grupos. Como se observó en el ejemplo univariante, los tamaños de los grupos son de 60 y 40, excediendo ambos los tamaños mínimos recomendados. Estos tamaños muestrales deben ser adecuados para proporcionar la potencia de 0,80 recomendada para efectos tamaño medianos.

Tercer paso: *Supuestos básicos del MANOVA*

Antes de calcular los contrastes estadísticos para las diferencias de las medias entre los grupos, el investigador primero debe determinar si las medidas dependientes están significativamente correlacionadas. El contraste que más ampliamente se emplea para este fin es el contraste de esfericidad de Barlett. Este contraste examina las correlaciones entre todas las variables dependientes y evalúa si conjuntamente existe intercorrelación significativa. En nuestro ejemplo, existe un grado de intercorrelación significativo (0,657) (véase Tabla 6.5).

El otro supuesto básico se refiere a la homogeneidad de las matrices de varianzas/covarianzas entre los dos grupos. El primer análisis valora la homogeneidad univariante de la varianza entre los dos grupos. Como se muestra en la Tabla 6.5, los contrastes univariantes para ambas variables no son significativos. El próximo paso es evaluar conjuntamente las variables dependientes contrastando la igualdad del total de las matrices de varianzas/covarianzas entre los grupos. Observando la Tabla 6.5, se detecta una diferencia en la correlación entre las dos variables dependientes en los dos grupos (0.823 para los clientes que utilizan la compra de especificación y 0,559 para aquellos que utilizan el análisis del valor total), que ilustra los diferentes niveles de covarianza. El contraste para la igualdad de las matrices de varianzas/covarianzas es el contraste M de Box, que en este ejemplo tiene un nivel de significación de 0,01. Dada la sensibilidad de este contraste, el nivel de significación se considera aceptable y se procede al análisis.

Cuarto paso: *Estimación del modelo MANOVA y valoración del ajuste global*

Las medias de cada variable dependiente (nivel de uso y satisfacción) para los dos grupos de empresas se presentan en la Tabla 6.4. Para realizar el contraste, otra vez especificamos nuestro nivel de significación (0,05 en este ejemplo) como el máximo error de Tipo I permisible. Para

TABLA 6.4. Uso y satisfacción de empresas con el método de compra (datos básicos para la T^2 de Hotelling o un MANOVA de dos grupos)

Método de compra								
Grupo 1: Especificación por separado			*Grupo 2: Análisis de valor total*					
Observación	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Observación*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Observación*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*
2	43,0	4,3	1	32,0	4,2	51	41,0	5,0
3	48,0	5,2	5	58,0	6,8	55	39,0	3,7
4	32,0	3,9	7	46,0	5,8	56	47,0	4,2
6	45,0	4,4	9	63,0	5,4	58	65,0	6,0
8	44,0	4,3	11	32,0	4,3	59	46,0	5,6
10	54,0	5,4	12	47,0	5,0	61	54,0	4,8
13	39,0	4,4	14	38,0	5,0	62	60,0	6,1
24	36,0	3,7	15	54,0	5,9	63	47,0	5,3
27	36,0	3,7	16	49,0	4,7	64	36,0	4,2
30	46,0	5,1	17	38,0	4,4	66	45,0	4,9
31	43,0	3,3	18	40,0	5,6	67	59,0	6,0
34	47,0	3,8	19	54,0	5,9	69	58,0	4,3
35	35,0	4,1	20	55,0	6,0	72	55,0	3,9
36	39,0	3,6	21	41,0	4,5	73	51,0	4,9
37	44,0	4,8	22	35,0	3,3	74	60,0	5,1
39	29,0	3,9	23	55,0	5,2	76	49,0	5,2
40	28,0	3,3	25	49,0	4,9	77	42,0	5,1
41	40,0	3,7	26	49,0	5,9	78	47,0	5,1
45	38,0	3,2	28	54,0	5,8	80	56,0	5,1
48	43,0	4,7	29	49,0	5,4	81	59,0	4,5
52	53,0	5,2	32	53,0	5,0	82	47,0	5,6
53	50,0	5,5	33	60,0	6,1	84	37,0	4,4
54	32,0	3,7	38	46,0	5,1	88	36,0	4,3
57	62,0	6,2	42	58,0	6,7	90	60,0	6,1
60	50,0	5,0	43	53,0	5,9	91	49,0	4,4
65	40,0	3,4	44	48,0	4,8	92	39,0	5,5
68	46,0	4,5	46	54,0	6,0	93	43,0	5,2
70	49,0	4,8	47	55,0	4,9	95	31,0	4,0
71	50,0	5,4	49	57,0	4,9	97	60,0	5,2
75	41,0	4,1	50	53,0	3,8	100	33,0	4,4
79	39,0	3,3						
83	41,0	4,1						
85	53,0	5,6						
86	43,0	3,7						
87	51,0	5,5						
89	34,0	4,0						
94	36,0	3,6						
96	25,0	3,4						
98	38,0	3,7						
99	42,0	4,3						
Media	42,100	4,295		48,767	5,088			
Varianza	60,653	0,612		77,405	0,569			
Tamaño muestral	40			60				

Nota: T^2 de Hotelling = 26,333

TABLA 6.5. Información para la diagnosis del MANOVA de dos grupos

CONTRASTE DE LOS SUPUESTOS: HOMOGENEIDAD DE LAS MATRICES VARIANZA-COVARIANZA

Matrices varianza-covarianza
(valores en paréntesis son correlaciones)

	Grupo 1: Especificación por separado		*Grupo 2: Análisis del valor total*	
	X_9	X_{10}	X_9	X_{10}
X_9: Nivel de uso	60,656		77,402	
X_{10}: Nivel de satisfacción	5,011	0,611	3,707	0,569
	(0,823)		(0,559)	

Test de diagnóstico

	X_9: *Nivel de uso*		X_{10}: *Nivel de satisfacción*		*Global*	
	Estadístico	*Significación*	*Estadístico*	*Significación*	*Estadístico*	*Significación*
Contrastes univariantes						
C de Cochran	0,561	0,396	0,518	0,803		
Box de Bartlett	0,677	0,411	0,060	0,807		
Contraste de Levene	1,377	0,243	0,323	0,571		
Contraste multivariante						
M de Box					11,684	0,010

CONTRASTE DE SUPUESTO: CORRELACIÓN DE VARIABLES DEPENDIENTES

	Estadístico	*Significación*
Contraste de esfericidad de Bartlett	54,474	0,000
Intercorrelación: X_9 frente a X_{10}	0,657	

determinar el valor de T^2_{crit}, buscamos en la distribución F con 2 y 97 grados de libertad. Con un F_{crit} de 3,09, la T^2_{crit} puede ser calculada como sigue

$$T^2_{crit} = \frac{p(N_1 + N_2 - 2)}{N_1 + N_2 - p - 1} \times F_{crit}$$

$$= \frac{2(60 + 40 - 2)}{60 + 40 - 2 - 1} \times 3{,}09$$

$$= 6{,}24$$

Como se muestra al final de la Tabla 5.3, el valor calculado de la T^2 de Hotelling es 26,33. Dado que este valor excede a la T^2_{crit}, rechazamos la hipótesis nula y concluimos que el método de compra ha tenido algún efecto sobre el conjunto de medidas dependientes. Además, la potencia del contraste multivariante fue casi de 1,0, lo que indica que los tamaños muestrales y el efecto tamaño fueron suficientes para asegurar que las diferencias significativas serían detectadas si éstas existieran más allá de las diferencias debidas meramente al error muestral.

Quinto paso: *Interpretación de los resultados*

Dada la significación del contraste multivariante al detectar las diferencias de los grupos sobre el valor teórico dependiente (vector de medias), el investigador debe ahora examinar los resultados para valorar su consistencia lógica. El grupo que utiliza el análisis del valor total no solamente tiene un mayor nivel de uso (véase Figura 6.6) sino que también tiene un mayor nivel de satisfacción

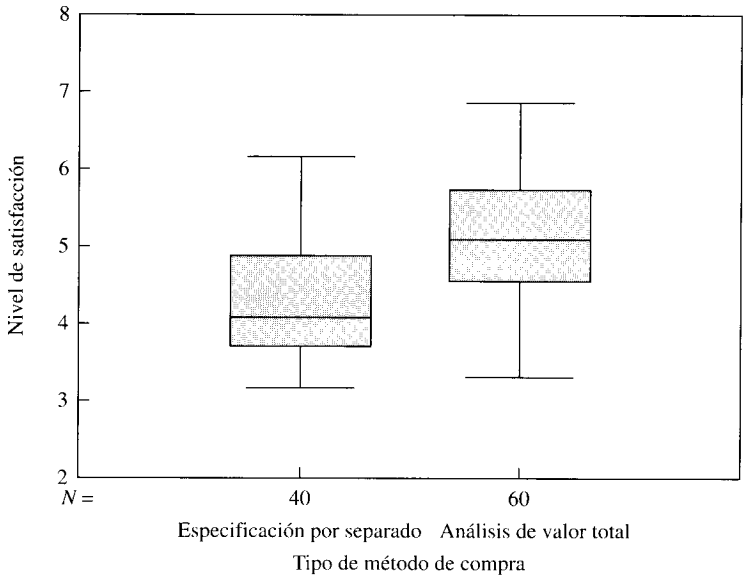


FIGURA 6.7. Dibujos de caja de la segunda variable dependiente, satisfacción, en un MANOVA de dos grupos.

(Figura 6.7). La cuestión que ahora debe evaluar el investigador es si ambas variables dependientes son significativamente diferentes o si los resultados obtenidos se deben principalmente a las diferencias solamente de una de las dos variables dependientes.

Un contraste *post hoc* de claro interés es si el método de compra tuvo un impacto sobre el nivel de uso (X_9) o sobre la satisfacción (X_{10}), considerando cada uno de forma separada. Las medias de los grupos y los contrastes univariantes de las diferencias de las medias son como siguen:

	Variables dependientes	
	X_9 *Nivel de uso*	X_{10} *Nivel de satisfacción*
Media del grupo		
Análisis de valor total	48,77	5,09
Compra de especificación	42,10	4,30
Diferencia	+6,67	+0,79
Prueba univariante de diferencias de grupo		
Estadístico *t*	3,96	5,08
Nivel de significación	0,000	0,000

El estadístico *t* para la diferencia en el nivel de uso se calculó anteriormente dando un resultado de 3,96, y para la diferencia en los niveles de satisfacción el estadístico *t* es de 5,08. Ambos estadísticos exceden la raíz cuadrada de la T^2_{crit} ($\sqrt{6,24} = 2,50$). Con esto podemos concluir que el método de compra tuvo un efecto positivo sobre el nivel de uso y satisfacción. En el empleo de la T^2, también estamos seguros de que la probabilidad del error de Tipo I es del 5 por ciento entre ambos contrastes post hoc.

Un segundo análisis del valor teórico dependiente es el contraste de reducción, con el que se examina la significación de las diferencias de los grupos pero permitiendo intercorrelación de la variable dependiente. La Tabla 6.6 refleja que tanto X_9 como X_{10} son significativamente diferentes, incluso cuando se tiene en cuenta su intercorrelación. De esta manera, después de examinar los resultados y los contrastes vistos anteriormente, el investigador puede concluir con seguridad que los dos grupos, tanto conjunta como individualmente, difieren significativamente en ambas variables.

TABLA 6.6. Contrastes de reducción en un MANOVA de dos grupos

CONTRASTES F DE REDUCCIÓN DE ROY BARGMANN						
				Grados de libertad		
Variable	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*F de reducción*	*entre grupos*	*intra grupos*	*Significación de la* F *de reducción*
X_9	1066,667	70,738	15,079	1	98	0,000
X_{10}	3,246	0,336	9,653	1	97	0,002

Ejemplo 2: Diferencia entre *k* grupos independientes

El diseño aleatorizado de dos grupos (Ejemplo 1) es un caso especial de un diseño aleatorizado de k grupos más general. En el caso general, cada encuestado es aleatoriamente asignado a uno de los k niveles (grupos) del tratamiento (variable independiente). En el caso univariante, se mide una única variable dependiente métrica, y la hipótesis nula es que todas las medias de los grupos son iguales (por ejemplo, $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_K$). En el caso multivariante, son medidas múltiples variables dependientes métricas, y la hipótesis nula es que las puntuaciones de los vectores de medias de los grupos son iguales (por ejemplo, $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$, donde μ representa a un vector o conjunto de puntuaciones medias). Para un análisis univariante, el contraste estadístico es el estadístico F resultante del ANOVA. En un análisis multivariante, examinaremos dos de los contrastes más ampliamente utilizados como son el estadístico mayor raíz característica (*mrc*) (también denominado mayor raíz de Roy) y el lambda de Wilks (también conocido como criterio del ratio de verosimilitud de Wilks o estadístico U).

Un enfoque univariante: ANOVA con *k* grupos

Primer paso: *Objetivos del ANOVA*

El estudio de HATCO también preguntó a los clientes para poder clasificar el tipo de compra que realizan con HATCO (X_{14}) como situación de nueva tarea, recompra modificada o recompra directa (véase Capítulo 1 para una discusión más detallada de estos datos). HATCO también está interesado en conocer si el uso varía entre el tipo de situación de compra. La hipótesis nula global que ahora HATCO desea contrastar es que $\mu_1 = \mu_2 = \mu_3$ (por ejemplo, los tres grupos son iguales en cuanto a su nivel de uso).

Segundo paso: *Diseño de la investigación del ANOVA*

En la muestra, 34 empresas indicaron que la mejor situación de compra que caracterizaba su relación con HATCO era la tarea nueva, 32 empresas respondieron que era la recompra modificada y 34 empresas eligieron la recompra directa. Estos tamaños muestrales son adecuados para obtener suficiente potencia con efectos tamaño medianos o grandes (véase Tabla 6.2). Si el efecto tamaño fuese pequeño y los tamaños muestrales fueran reducidos debidos a datos ausentes u otros factores, la potencia caería por debajo de los niveles recomendados. Entonces el investigador necesitaría evaluar cuidadosamente los contrastes estadísticos para la potencia y la significación práctica de las diferencias.

Tercer paso: *Supuestos básicos del ANOVA*

Los contrastes univariantes para la homogeneidad de la varianza de X_9, nivel de uso, entre los tres grupos no muestran diferencias significativas con cualquiera de las tres tareas. De esta forma, otra vez el investigador puede dejar de considerar, como se hizo en el ejemplo de dos grupos, los diversos tamaños de los grupos como influyentes sobre los contrastes estadísticos para las diferencias de los grupos.

Cuarto paso: *Estimación del modelo ANOVA y valoración del ajuste global*

El modelo ANOVA contrasta las diferencias en las medias de los grupos entre aquellos clientes de HATCO que utilizan uno de los tres métodos de compra. La Figura 6.8 incluye una representación gráfica de los encuestados por grupo, y los niveles de uso por grupo de las 100 empresas son mostrados en la Tabla 6.7. Para llevar a cabo el contraste a mano, especificamos 0,05 como el porcentaje del error de Tipo I. Para determinar el valor de F_{crit}, buscamos en la distribución F con $(3 - 1) = 2$ y $(100 - 3) = 97$ grados de libertad con $\alpha = 0{,}05$. Encontramos que $F_{crit} = 3{,}09$. El cálculo del estadístico F del ANOVA está resumido en una tabla ANOVA similar a lo que se muestra en la Tabla 6.8. Los valores cuadrados medios tanto de las varianzas entre grupos como intra grupos se calculan como la suma de cuadrados (suma de las desviaciones al cuadrado) divididos por adecuados grados de libertad. Como se observa en la Tabla 6.8, se obtiene un estadístico $F = 106{,}66$ (2749,833 / 25,776). Dado que este valor excede al F_{crit}, podemos concluir que las medias de todos los grupos no son iguales.

Quinto paso: *Interpretación de los resultados*

Como se muestra en la Tabla 6.7, las medias de los grupos para el nivel de uso son nueva tarea (36,91), recompra modificada (46,53) y recompra directa (54,88).

Observando estas medias, notamos que el uso de HATCO se incrementa a medida que vamos de nueva tarea a recompra modificada y a recompra directa. Una hipótesis de interés es si existe diferencia significativa entre la nueva tarea o la recompra modificada frente a la recompra direc-

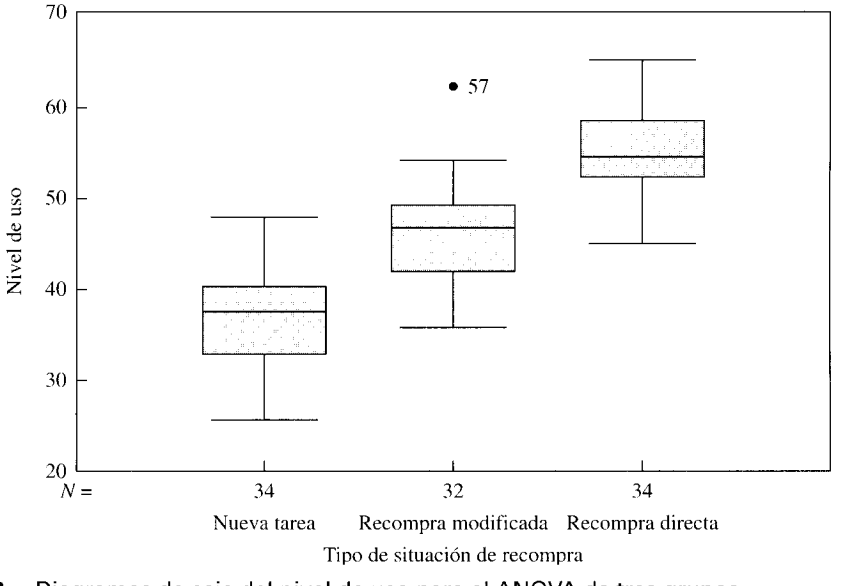


FIGURA 6.8. Diagramas de caja del nivel de uso para el ANOVA de tres grupos.

TABLA 6.7. Nivel de uso de empresas según el tipo de situación de la compra (datos básicos de un ANOVA de tres grupos)

Situación de compra					
Grupo 1: Nueva tarea		*Grupo 2: Recompra modificada*		*Grupo 3: Recompra directa*	
Empresa	X_9 *(Nivel de uso)*	*Empresa*	X_9 *(Nivel de uso)*	*Empresa*	X_9 *(Nivel de uso)*
1	32,0	3	48,0	5	58,0
2	43,0	6	45,0	9	63,0
4	32,0	8	44,0	15	54,0
7	46,0	10	54,0	16	49,0
11	32,0	12	47,0	19	54,0
13	39,0	17	38,0	20	55,0
14	38,0	18	40,0	23	55,0
22	35,0	21	41,0	26	49,0
24	36,0	25	49,0	28	54,0
27	36,0	30	46,0	29	49,0
31	43,0	37	44,0	32	53,0
34	47,0	44	48,0	33	60,0
35	35,0	48	43,0	38	46,0
36	39,0	51	41,0	42	58,0
39	29,0	52	53,0	43	53,0
40	28,0	53	50,0	46	54,0
41	40,0	56	47,0	47	55,0
45	38,0	57	62,0	49	57,0
54	32,0	60	50,0	50	53,0
55	39,0	64	36,0	58	65,0
65	40,0	66	45,0	59	46,0
75	41,0	68	46,0	61	54,0
79	39,0	70	49,0	62	60,0
83	41,0	71	50,0	63	47,0
84	37,0	76	49,0	67	59,0
86	43,0	77	42,0	69	58,0
88	36,0	82	47,0	72	55,0
89	34,0	85	53,0	73	51,0
94	36,0	87	51,0	74	60,0
95	31,0	91	49,0	78	47,0
96	25,0	92	39,0	80	56,0
98	38,0	93	43,0	81	59,0
99	42,0			90	60,0
100	33,0			97	60,0
Media	36,912		46,531		54,882
Varianzas	25,593		28,132		23,746
Tamaño muestral	34		32		34

TABLA 6.8. Resultados del ANOVA de tres grupos: Nivel de uso de la situación de la compra

Varianza	*Suma de cuadrados*	*Media al cuadrado*	*Grados de libertad*	*Ratio F*
Entre grupos	5498,767	2749,383	2	106,666
Error dentro de los grupos	2500,233	25,776	97	

ta [por ejemplo, (36,91 + 46,53) / 2 frente a 54,88]. Este tipo de situación puede ser contrastada con uno de los procedimientos a priori. El contraste es significativo (suponiendo $\alpha = 0,05$; las expresiones para el cálculo se pueden encontrar en textos más orientados a los estadísticos). Así podemos concluir que la situación de recompra directa presenta unos mayores niveles de uso que las otras dos situaciones de compra. Otro enfoque es utilizar uno de los procedimientos post hoc que contraste las diferencias de todos los grupos e identifique aquellas diferencias que son estadísticamente significativas. Uno de estos es el método de Scheffé, quizá el más utilizado de los métodos post hoc. En este ejemplo, a medida que controlamos el porcentaje del error total para que no exceda 0,05, el contraste de Scheffe identifica que todos los grupos son estadísticamente diferentes. A partir de esto, el investigador conocerá que las diferencias significativas son debidas a comparaciones en cada grupo y que no son diferencias específicas entre algunos grupos solamente.

En resumen, el ANOVA univariante sugiere que el tipo de situación de la compra lleva a mayores niveles de uso. Los contrastes post hoc posibilitan al investigador que identifique estas diferencias significativas más fácilmente y que mantenga el control estadístico sobre el nivel de significación total.

Un enfoque multivariante: MANOVA con *k* grupos

En los diseños de k grupos donde se miden múltiples variables dependientes, muchos investigadores llevan a cabo una serie de contrastes F individuales (ANOVAs) hasta que todas las variables dependientes han sido analizadas. Como debe sospechar el lector, este enfoque sufre de las mismas deficiencias que se dan cuando se utilizan una serie de contrastes t entre múltiples variables dependientes; es decir, un conjunto de contrastes F con el ANOVA (1) provoca un incremento del porcentaje del error de Tipo I, y (2) ignora la posibilidad de que algunas combinaciones de las variables dependientes pueden proporcionar una clara evidencia de las diferencias del total de los grupos. Además, dado que los contrastes F individuales ignoran las correlaciones entre las variables independientes, éstos emplean menos información de la disponible para evaluar las diferencias del total de los grupos.

Una vez más el MANOVA proporciona una solución a estos problemas. El MANOVA resuelve el problema del porcentaje del error de Tipo I proporcionando un único contraste global de las diferencias de los grupos con un nivel α determinado. Resuelve el problema de la variable compuesta formando implícitamente y contrastando la(s) combinación(es) lineal(es) de las variables dependientes que aportan la evidencia más fuerte de las diferencias del total de los grupos.

Primer paso: *Objetivos del MANOVA*

En nuestro ejemplo univariante anterior, HATCO valoró su actuación entre las empresas que tienen uno de las tres situaciones de compra (X_9) (compra de nueva tarea, recompra modificada o recompra directa). Para convertir este ejemplo en un ejemplo multivariante, se requieren al menos dos variables dependientes. Al igual que en nuestra anterior extensión multivariante de un ejemplo univariante, supongamos que HATCO también desea examinar las diferencias en satisfacción (X_{10}) con HATCO entre los tres grupos (véase Tabla 6.9).

La hipótesis nula que ahora HATCO desea contrastar es que los tres vectores simples de las puntuaciones medias son iguales.

Segundo paso: *Diseño de la investigación del MANOVA*

Como se trató en el ejemplo univariante del análisis de tres grupos, los tamaños muestrales son los adecuados en relación con el número de variables dependientes. Una cuestión más importante es el efecto de los tamaños muestrales de los grupos sobre la potencia estadística de los contrastes de

TABLA 6.9. Nivel de uso y nivel de satisfacción por la situación de compra (datos básicos para un MANOVA de tres grupos)

Situación de compra								
Grupo 1: Nueva tarea			*Grupo 2: Recompra modificada*			*Grupo 3: Recompra directa*		
Empresa	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Empresa*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Empresa*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*
1	32,0	4,2	3	48,0	5,2	5	58,0	6,8
2	43,0	4,3	6	45,0	4,4	9	63,0	5,4
4	32,0	3,9	8	44,0	4,3	15	54,0	5,9
7	46,0	5,8	10	54,0	5,4	16	49,0	4,7
11	32,0	4,3	12	47,0	5,0	19	54,0	5,9
13	39,0	4,4	17	38,0	4,4	20	55,0	6,0
14	38,0	5,0	18	40,0	5,6	23	55,0	5,2
22	35,0	3,3	21	41,0	4,5	26	49,0	5,9
24	36,0	3,7	25	49,0	4,9	28	54,0	5,8
27	36,0	3,7	30	46,0	5,1	29	49,0	5,4
31	43,0	3,3	37	44,0	4,8	32	53,0	5,0
34	47,0	3,8	44	48,0	4,8	33	60,0	6,1
35	35,0	4,1	48	43,0	4,7	38	46,0	5,1
36	39,0	3,6	51	41,0	5,0	42	58,0	6,7
39	29,0	3,9	52	53,0	5,2	43	53,0	5,9
40	28,0	3,3	53	50,0	5,5	46	54,0	6,0
41	40,0	3,7	56	47,0	4,2	47	55,0	4,9
45	38,0	3,2	57	62,0	6,2	49	57,0	4,9
54	32,0	3,7	60	50,0	5,0	50	53,0	3,8
55	39,0	3,7	64	36,0	4,2	58	65,0	6,0
65	40,0	3,4	66	45,0	4,9	59	46,0	5,6
75	41,0	4,1	68	46,0	4,5	61	54,0	4,8
79	39,0	3,3	70	49,0	4,8	62	60,0	6,1
83	41,0	4,1	71	50,0	5,4	63	47,0	5,3
84	37,0	4,4	76	49,0	5,2	67	59,0	6,0
86	43,0	3,7	77	42,0	5,1	69	58,0	4,3
88	36,0	4,3	82	47,0	5,6	72	55,0	3,9
89	34,0	4,0	85	53,0	5,6	73	51,0	4,9
94	36,0	3,6	87	51,0	5,5	74	60,0	5,1
95	31,0	4,0	91	49,0	4,4	78	47,0	5,1
96	25,0	3,4	92	39,0	5,5	80	56,0	5,1
98	38,0	3,7	93	43,0	5,2	81	59,0	4,5
99	42,0	4,3				90	60,0	6,1
100	33,0	4,4				97	60,0	5,2
Media	36,912	3,929		46,531	5,003		54,882	5,394
Varianza	25,593	0,282		28,132	0,237		23,746	0,508
Tamaño muestral	34			32			34	

las diferencias entre los grupos. Atendiendo a la Tabla 6.1, los tamaños muestrales de 30 y superiores proporcionarán una potencia adecuada para efectos tamaño grandes y de alguna manera bajos niveles de potencia para efectos tamaño medianos. Sin embargo, estos tamaños muestrales, no son adecuados para proporcionar la potencia de 0,80 recomendada para efectos tamaño pequeños. El tamaño muestral requerido para efectos tamaño pequeños en esta situación sería de 98 encuestados por grupo. Por ello, cualquier resultado no significativo deberá ser examinado cuidadosamente para evaluar si el efecto tamaño tiene una importancia primordial, ya que la baja potencia estadística evita que se consideren como estadísticamente significativos.

Tercer paso: *Supuestos básicos del MANOVA*

Los dos contrastes univariantes para la homogeneidad de la varianza indican una diferencia no significativa para X_9, el nivel de uso, y se obtienen resultados contradictorios para X_{10}, la satisfacción con HATCO (Tabla 6.10). En el caso de la satisfacción con HATCO, la combinación de los dos contrastes (0,033 y 0,070) proporciona un nivel suficiente de no significación para el contraste de homogeneidad de la varianza para llevar a cabo el contraste multivariante. Empleando el contraste *M* de Box para la homogeneidad de las matrices de varianzas/covarianzas, encontramos que los grupos no presentan diferencias significativas.

El segundo supuesto a contrastar es la correlación entre las variables dependientes. En este caso el contraste de Bartlett de esfericidad tiene un nivel de significación de 0,002, satisfaciendo por tanto el nivel necesario de intercorrelación para justificar el uso del MANOVA. Véase Tabla 6.10 para más detalles.

Cuarto paso: *Estimación del modelo MANOVA y valoración del ajuste global*

Del examen de los diagramas de caja de los encuestados en cada grupo para el nivel de uso (véase Figura 6.8) y satisfacción (Figura 6.9), se observa que ambas variables pueden diferir entre los tres grupos. El objetivo del contraste multivariante es valorar estas diferencias conjuntamente en lugar de individualmente con contrastes univariantes. La Tabla 6.11 refleja un resumen del resultado del MANOVA realizado con el conjunto de datos de la Tabla 6.9. El criterio de Pillai tiene un nivel de significación (0,0000) muy por debajo de nuestro nivel preespecificado de 0,05. El valor

TABLA 6.10. Información para la diagnosis en el MANOVA de tres grupos

CONTRASTE DE LOS SUPUESTOS: HOMOGENEIDAD DE LAS MATRICES VARIANZA-COVARIANZA

Matrices varianza-covarianza (valores en paréntesis son correlaciones)

	Grupo 1: Nueva tarea		*Grupo 2: Recompra modificada*		*Grupo 3: Recompra directa*	
	X_9	X_{10}	X_9	X_{10}	X_9	X_{10}
X_9: Nivel de uso	25,598		28,128		23,743	
X_{10}: Nivel de satisfacción	0,648	0,282	1,366	0,237	0,763	0,509
	(0,241)		(0,529)		(0,219)	

Contrastes diagnósticos

	X_9: *Nivel de uso*		X_{10}: *Nivel de satisfacción*		*Global*	
	Estadístico	*Significación*	*Estadístico*	*Significación*	*Estadístico*	*Significación*
Contrastes univariantes						
C de Cochran	0,363	0,965	0,495	0,033		
Box de Bartlett	0,114	0,892	2,670	0,070		
Contraste de Levene	0,056	0,945	3,302	0,041		
Contraste multivariante						
M de Box					9,796	0,147

CONTRASTE DE SUPUESTO: CORRELACIÓN DE VARIABLES DEPENDIENTES

	Estadístico	*Significación*
Contraste de esfericidad de Bartlett	9,480	0,002
Intercorrelación: X_9 frente a X_{10}	0,307	

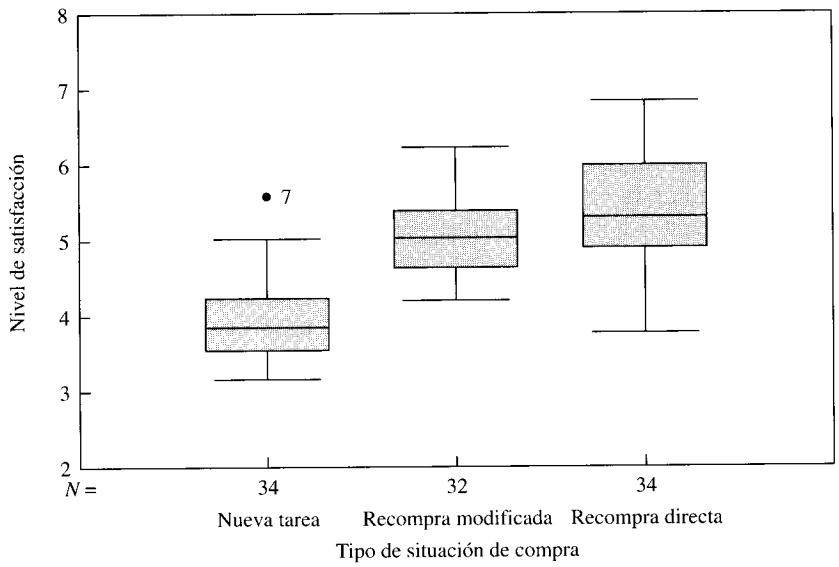


FIGURA 6.9. Dibujos de caja de la segunda variable dependiente, la satisfacción, en un MANOVA de tres grupos.

de la mayor raíz característica es 0,723. Atendiendo a la distribución de *mrc* con adecuados grados de libertad y poniendo un $\alpha = 0{,}05$, encontramos que $mrc_{crit} = 0{,}310$. Dado que 0,723 excede este valor, podemos concluir una vez más que los vectores de medias de los tres grupos no son iguales. Como se ve en la Tabla 6.11, el programa estadístico estima un nivel de significación de 0,000 para esta medida también. El valor del lambda de Wilks, (véase Tabla 6.11), es 0,264. Un estadístico F aproximado asociado a este valor del lambda de Wilks es 45,4. Con 4 y 192 grados de libertad y un nivel α de 0,05, la $F_{crit} = 2{,}41$. Como 45,4 excede en mucho este valor, otra vez obtenemos la misma conclusión de que los vectores de las medias de los tres grupos no son iguales. Empleando cualquiera de las medidas de las diferencias multivariantes se obtiene la misma conclusión: las variables dependientes combinadas, el uso y la satisfacción, varían entre las tres situaciones de compra.

Quinto paso: *Interpretación de los resultados*

Con la interpretación de los resultados (mostrados en la Tabla 6.9), el investigador tiene que examinar primero las medias de los grupos. Ya hemos visto que la pauta para el nivel de uso incrementando a través de los tres grupos, y la media de grupo para el nivel de satisfacción siguen la misma pauta general: nueva tarea (3,93), recompra modificada (5,00) y recompra simple (5,39).

Puede resultar de interés realizar un conjunto de contrastes post hoc. Por ejemplo, ¿las diferencias de las medias de los grupos son estadísticamente significativas si consideramos cada variable dependiente de forma aislada? Podemos estudiar esta situación con los contrastes F individuales para las dos variables dependientes. Al igual que se obtuvo en los resultados multivariantes, ambas variables muestran diferencias significativas entre los grupos. El análisis de reducción, como se muestra en la Tabla 6.11, refleja que ambas variables tienen diferencias únicas entre los grupos; es decir, no están tan altamente correlacionadas como para que no hubiera diferencias únicas en la satisfacción después de que se hubieran tenido en cuenta los efectos del uso. Este resultado implica que la situación de la compra tiene efectos separados significativos sobre la satisfacción que no están relacionados con un nivel de uso de la empresa. Finalmente, ¿se encuentran diferencias en el uso y en la satisfacción entre todas las situaciones de compra? Por ejem-

TABLA 6.11. Tabla resumen del MANOVA de tres grupos

CONTRASTES DE SIGNIFICACIÓN

Nombre del contraste	*Valor*	*F aproximada*	*Grados de libertad*		*Significación del estadístico F*
			entre grupos	*intra grupos*	
Criterio de Pillai	0,771	30,419	4	194	0,000
Traza de Hotelling	2,655	63,052	4	190	0,000
Lambda de Wilks	0,264	45,411	4	192	0,000
Mrc de Roy	0,723				

POTENCIA ESTADÍSTICA DE LOS TESTS DEL MANOVA

	Efecto tamaño	*Potencia*
Criterio de Pillai	0,385	1,000
Traza de Hotelling	0,570	1,000
Lambda de Wilks	0,486	1,000

Tests *f* univariantes

Variable	*Suma de cuadrado entre grupos*	*Suma de cuadrado intra grupos*	*Grados de libertad*	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*Estadístico F*	*Significación*
X_9	5498,767	2500,233	2 y 97	2749,383	25,776	106,666	0,000
X_{10}	39,007	33,459	2 y 97	19,503	0,345	56,542	0,000

CONTRASTES DE RECUCCIÓN DE ROY BARGMAN

Variable	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*Reducción F*	*Grados de libertad*		*Significación de la reduccción F*
				entre grupos	*intra grupos*	
X_9	2749,383	25,776	106,666	2	97	0,000
X_{10}	2,783	0,316	8,819	2	96	0,000

plo, ¿las empresas con situación de recompra modificada presentan una diferencia en la satisfacción cuando son comparadas con las de recompra directa? Se puede responder a todas estas cuestiones con la extensión multivariante de los procedimientos de contraste señalados anteriormente. El estudio de las funciones discriminantes (que tienen que ser obtenidas empleando un programa de análisis discriminante, como el descrito en el Capítulo 5) son generalmente más útiles en las primeras etapas del análisis post hoc a medida que se incrementa el número de variables dependientes.

Ejemplo 3: Diseño factorial del MANOVA con dos variables independientes

En los anteriores dos casos, los análisis MANOVA han sido extensiones de análisis univariantes de dos y tres grupos. En este ejemplo, abordaremos un diseño factorial multivariante —el empleo de dos variables independientes como tratamientos para analizar las diferencias de dos variables dependientes. A lo largo de nuestra discusión, evaluaremos los efectos conjuntos o interactivos entre los dos tratamientos sobre las variables dependientes tanto separada como conjuntamente.

Primer paso: Objetivos del MANOVA

En las cuestiones de investigación anteriores, HATCO ha considerado solamente los efectos de una sola variable tratamiento sobre las variables dependientes. Pero también se debe considerar la posibilidad de efectos conjuntos entre dos o más variables independientes. Después de una deliberación, se ha propuesto el análisis centrándose en la extensión del análisis de tres grupos anterior para captar las diferencias en las dos variables dependientes, nivel de uso (X_9) y de satisfacción (X_{10}), pero considerando ahora no sólo los efectos de las situaciones de la compra (X_{14}) sino también el tipo de industria (X_{13}). El objetivo es reducir la varianza dentro del grupo de los grupos de la situación de compra introduciendo un segundo tratamiento —el tipo de industria. Éste creará un grupo separado para cada tipo de industria dentro de cada situación de compra. Los diagramas de caja de la Figura 6.10 muestran que los dos tipos de industria no varían sustancialmente en ninguna de las variables dependientes. Pero esto no excluye su utilidad como un medio de reducir la varianza dentro del grupo. Los datos de cada grupo (combinación de situación de compra y tipo de industria) aparecen en la Tabla 6.12. De esta manera, el efecto del tipo de industria puede ser evaluado simultáneamente con la situación de la compra junto con el examen de la posible existencia de efectos de interacción.

Segundo paso: Diseño de investigación del MANOVA

En el diseño factorial de dos variables independientes —X_{14} y X_{13}— se plantea la cuestión del tamaño muestral adecuado en los diferentes grupos. Ya que existen tres niveles de X_{14} (tarea nueva, recompra modificada y recompra directa) y dos niveles de X_{13} (categoría uno SIC y categoría dos SIC), este es un diseño 3 × 2 con seis grupos. El investigador debe asegurarse al crear el diseño factorial, que cada grupo tiene suficiente tamaño muestral para (1) cumplir los requisitos mínimos de tamaño de grupos que excedan el número de variables dependientes, y también para (2) proporcionar la potencia estadística para evaluar las diferencias que se consideren prácticamente significativas. En este caso, los tamaños muestrales varían entre 16 y 18 encuestados por grupo. Esto excede el número de variables dependientes (dos) en cada grupo, pero la potencia estadística es bastante baja. Aunque no se hayan tabulado valores para un diseño factorial, los valores para un MA-

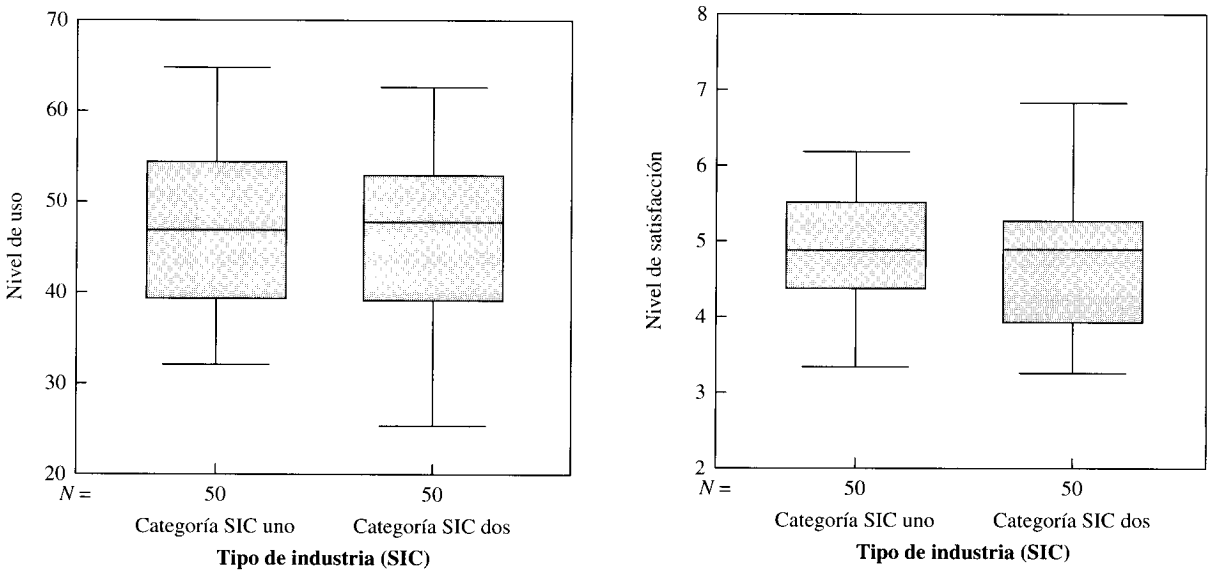


FIGURA 6.10. Diagramas en caja de nivel de uso y satisfacción para la segunda variable independiente, tipo de industria, en diseño factorial de MANOVA 3 × 2.

TABLA 6.12. Uso y nivel de satisfacción para la situación de la compra y el tipo de industria (datos básicos para un diseño factorial (MANOVA 3 × 2)

Categoría SIC uno

Grupo 1: Puntuaciones de nueva tarea			*Grupo 3: Análisis de valor total*			*Grupo 5: Puntuaciones de recompra directa*		
Empresa	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Empresa*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Empresa*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*
2	43,0	4,3	8	44,0	4,3	15	54,0	5,9
11	32,0	4,3	10	54,0	5,4	16	49,0	4,7
13	39,0	4,4	18	40,0	5,6	20	55,0	6,0
22	35,0	3,3	21	41,0	4,5	23	55,0	5,2
24	36,0	3,7	25	49,0	4,9	32	53,0	5,0
27	36,0	3,7	30	46,0	5,1	33	60,0	6,1
31	43,0	3,3	44	48,0	4,8	43	53,0	5,9
34	47,0	3,8	51	41,0	5,0	46	54,0	6,0
35	35,0	4,1	53	50,0	5,5	58	65,0	6,0
55	39,0	3,7	57	62,0	6,2	62	60,0	6,1
75	41,0	4,1	66	45,0	4,9	67	59,0	6,0
83	41,0	4,1	68	46,0	4,5	69	58,0	4,3
84	37,0	4,4	77	42,0	5,1	73	51,0	4,9
88	36,0	4,3	85	53,0	5,6	80	56,0	5,1
94	36,0	3,6	87	51,0	5,5	81	59,0	4,5
98	38,0	3,7	92	39,0	5,5	90	60,0	6,1
99	42,0	4,3						
100	33,0	4,4						
Media	38,278	3,972		46,937	5,150		56,313	5,488
Varianza	15,390	0,141		37,663	0,252		16,761	0,423
Tamaño muestral	18			16			16	

Categoría SIC dos

Grupo 2: Puntuaciones de nueva tarea			*Grupo 4: Análisis de valor total*			*Grupo 6: Puntuaciones de recompra directa*		
Empresa	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Empresa*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*	*Empresa*	X_9 *(Nivel de uso)*	X_{10} *(Nivel de satisfacción)*
1	32,0	4,2	3	48,0	5,2	5	58,0	6,8
4	32,0	3,9	6	45,0	4,4	9	63,0	5,4
7	46,0	5,8	12	47,0	5,0	19	54,0	5,9
14	38,0	5,0	17	38,0	4,4	26	49,0	5,9
36	39,0	3,6	37	44,0	4,8	28	54,0	5,8
39	29,0	3,9	48	43,0	4,7	29	49,0	5,4
40	28,0	3,3	52	53,0	5,2	38	46,0	5,1
41	40,0	3,7	56	47,0	4,2	42	58,0	6,7
45	38,0	3,2	60	50,0	5,0	47	55,0	4,9
54	32,0	3,7	64	36,0	4,2	49	57,0	4,9
65	40,0	3,4	70	49,0	4,8	50	53,0	3,8
79	39,0	3,3	71	50,0	5,4	59	46,0	5,6
86	43,0	3,7	76	49,0	5,2	61	54,0	4,8
89	34,0	4,0	82	47,0	5,6	63	47,0	5,3
95	31,0	4,0	91	49,0	4,4	72	55,0	3,9
96	25,0	3,4	93	43,0	5,2	74	60,0	5,1
						78	47,0	5,1
						97	60,0	5,2
Media	35,375	3,881		46,125	4,856		53,611	5,311
Varianza	34,117	0,456		20,115	0,192		27,668	0,601
Tamaño muestral	16			16			18	

NOVA de seis grupos señalan que este tamaño muestral detectará mejor solamente efectos tamaño moderadamente grandes o con una potencia de 0,70 [15]. De esta manera, el investigador tiene que reconocer que a menos que los efectos tamaño sean importantes, el tamaño muestral limitado de cada grupo impide la identificación de diferencias significativas.

Tercer paso: Supuestos básicos del MANOVA

El supuesto de mayor importancia, al igual que ocurría en los análisis MANOVA anteriores, es el de la homogeneidad de las matrices de varianzas/covarianzas entre los grupos. En este ejemplo, se tienen seis grupos para contrastar el supuesto. Los contrastes univariantes para el nivel de uso y satisfacción no son significativos, excepto el contraste de Bartlett-Box para la satisfacción, que tiene un nivel de significación de 0,038 (véase Tabla 6.13). Observando que los contrastes univariantes no son significativos, el investigador puede continuar con el contraste multivariante. El contraste *M* de Box presenta un nivel de significación de 0,09, por lo que nos permite aceptar la hipótesis nula de homogeneidad de las matrices de varianzas/covarianzas a un nivel de significación de 0,05. Al cumplirse este supuesto podemos interpretar directamente los resultados sin tener en consideración los tamaños de los grupos, el nivel de las covariaciones en los grupos y así sucesivamente.

El segundo supuesto es la correlación de las medidas dependientes, la cual es evaluada con el contraste de Bartlett de esfericidad. En este ejemplo, la significación es de 0,004, indicativo de un nivel de significación de correlación entre las dos medidas dependientes (véase Tabla 6.13).

Cuarto paso: Estimación del modelo MANOVA y valoración del ajuste global

Los diagramas de caja para cada variable dependiente entre los seis grupos (Figura 6.11) muestran que sí parecen existir diferencias entre ellos. Las diferencias más claras se dan entre las situaciones de compra, pero también se observan dentro de cada situación de compra para los dos tipos de industria. El modelo MANOVA no solamente contrasta los efectos principales de ambas variables independientes sino también su efecto conjunto o interacción en las dos variables dependientes. El primer paso es examinar el efecto interacción y determinar si es estadísticamente significativo.

La Tabla 6.14 contiene los resultados del MANOVA para el contraste del efecto interacción. Los cuatro contrastes multivariantes indican que el efecto interacción no es significativo. Esto significa que las diferencias entre los tipos de industrias son aproximadamente iguales entre las tres situaciones de compra para las dos variables dependientes conjuntamente. Los contrastes univariantes confirman que estos resultados se dan para cada variable de forma separada. La Figura 6.12 refleja la carencia del efecto interacción para cada variable dependiente. En los gráficos de cada variable dependiente, las diferencias entre los dos tipos de industria son relativamente iguales entre las tres situaciones de compra. Al no existir un efecto interacción significativo, los efectos directos se pueden interpretar directamente sin ajustes.

Las Tablas 6.15 y 6.16 contienen los resultados del MANOVA para los efectos principales de la situación de compra y del tipo de industria. El tipo de industria (X_{13}) tiene un nivel de significación de 0,069 para los contrastes multivariantes, indicando una diferencia no significativa que se puede atribuir al tipo de industria. Sin embargo, el investigador debe considerar aumentar el nivel de significación requerido, ya que la potencia de los contrastes multivariantes se reduce debido a tamaños muestrales más pequeños por grupo. Si se realizase esto, entonces el tipo de industria sería considerado un efecto significativo. La segunda variable independiente, la satisfacción (X_{14}), muestra efectos altamente significativos en todos los contrastes multivariantes. En todos los casos, el nivel de significación excede 0,000. Además, la potencia estadística es 1,0, lo que indica que efectos tamaño muy grandes aseguran altos niveles de potencia incluso con tamaños muestrales pe-

TABLA 6.13. Información para la diagnosis del diseño factorial MANOVA 3 × 2

CONTRASTE DE LOS SUPUESTOS: HOMOGENEIDAD DE LAS MATRICES VARIANZA-COVARIANZA

Matrices varianza-covarianza
(valores en paréntesis son correlaciones)

	Categoría SIC uno					
	Grupo 1: Nueva tarea		*Grupo 3: Recompra modificada*		*Grupo 5: Recompra directa*	
	X_9	X_{10}	X_9	X_{10}	X_9	X_{10}
X_9: Nivel de uso	15,389		37,662		16,762	
X_{10}: Nivel de satisfacción	−0,162	0,141	1,643	0,252	0,984	0,423
	(−0,110)		(0,533)		(0,370)	
	Categoría SIC dos					
	Grupo 2: Nueva tarea		*Grupo 4: Recompra modificada*		*Grupo 6: Recompra directa*	
	X_9	X_{10}	X_9	X_{10}	X_9	X_{10}
X_9: Nivel de uso	34,117		20,117		27,663	
X_{10}: Nivel de satisfacción	1,461	0,456	1,053	0,192	0,375	0,600
	(0,370)		(0,536)		(0,092)	

Diagnóstico de contrastes

	X_9: Nivel de uso		X_{10}: Nivel deatisfacción		Global	
	Estadístico	*Significación*	*Estadístico*	*Significación*	*Estadístico*	*Significación*
Contrastes univariantes						
C de Cochran	0,248	0,447	0,291	0,115		
Box de Bartlett	1,062	0,380	2,363	0,038		
Contraste de Levene	1,332	0,257	1,519	0,191		
Contraste multivariante *M* de Box					24,050	0,090

CONTRASTE DEL SUPUESTO: CORRELACIÓN DE VARIABLES DEPENDIENTES

	Estadístico	*Significación*
Contraste de esfericidad de Bartlett	8,225	0,004
Intercorrelación: X_9 frente a X_{10}	0,292	

queños en los grupos. Se puede comparar el impacto de las dos variables independientes examinando el efecto tamaño relativo. El efecto tamaño para la situación de compra es de 8 a 10 veces más grande que el asociado con el nivel de uso. Esta comparación ofrece al investigador una evaluación de la significación práctica distinta a la de los contrastes de significación estadística. En este ejemplo, la situación de compra es el efecto dominante, y el tipo de industria tiene un débil efecto. Además, la interacción o efectos conjuntos entre los dos tratamientos no son significativos para ambas variables dependientes.

Quinto paso: Interpretación de los resultados

No se debe realizar una comparación entre los seis grupos con los contrastes post hoc, sino utilizar el método Scheffe, con el que se pueden llevar a cabo todas las comparaciones, a la vez que se controla el error de Tipo I. En este ejemplo, los tamaños muestrales relativamente pequeños y el gran

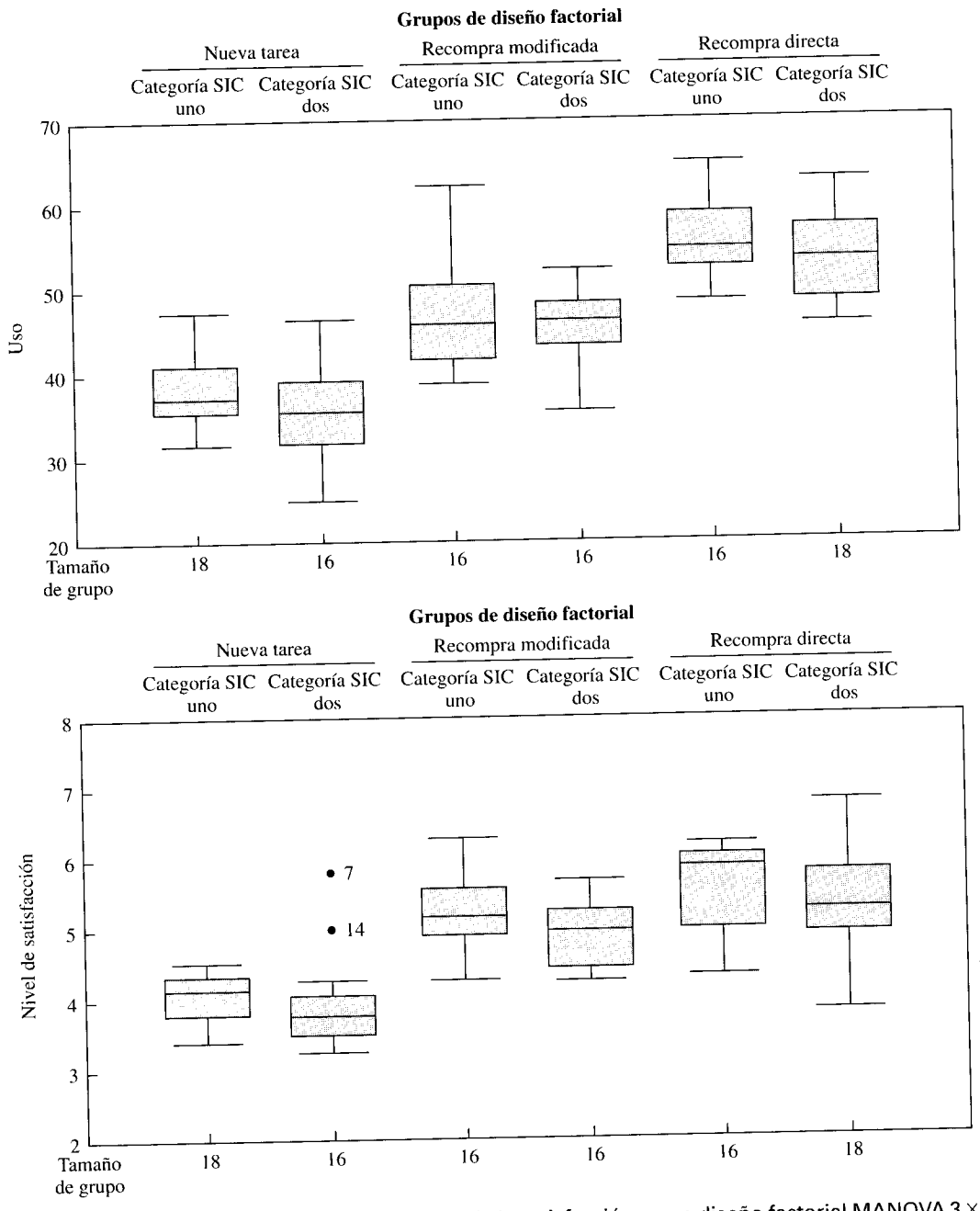


Figura 6.11. Dibujo en cajas del nivel de uso y de la satisfacción en un diseño factorial MANOVA 3 × 2.

número de contrastes requeridos para todas las comparaciones de los seis grupos producen unos niveles bajos de potencia estadística que solamente pueden detectar adecuadamente efectos tamaño muy grandes. El investigador debe por ello examinar las diferencias para la significación práctica junto con la significación estadística. Si se pueden realizar comparaciones específicas entre los grupos, entonces se pueden especificar comparaciones planificadas y contrastarlas directamente en el análisis.

Aunque las comparaciones entre los grupos están limitadas debido a los tamaños muestrales pequeños, aún se puede garantizar el examen de cada variable dependiente. Por ejemplo, se obtuvo que el tipo de industria (X_{13}) era no significativo a un nivel alfa de 0,05 cuando se evaluó el conjunto de variables dependientes. Pero si el investigador considera los contrastes univariantes, se

TABLE 6.14. Tabla resumen del diseño factorial MANOVA 3×2: efecto interacción

EFECTO INTERACCIÓN: TIPO DE INDUSTRIA (X_{13}) CON LA SITUACIÓN DE COMPRA (X_{14})

CONTRASTES MULTIVARIANTES DE SIGNIFICACIÓN

			Grados de libertad		
Nombre de contraste	*Valor*	*F Aproximada*	*entre grupos*	*intra grupos*	*Significación del estadístico F*
Criterio de Pillai	0,020	0,464	4	188	0,762
Traza de Hotelling	0,020	0,458	4	184	0,766
Lambda de Wilks	0,980	0,461	4	186	0,764
Mrc de Roy	0,019				

POTENCIA ESTADÍSTICA DE LOS TEST DEL MANOVA

	Efecto tamaño	*Potencia*
Criterio de Pillai	0,010	0,16
Traza de Hotelling	0,010	0,16
Lambda de Wilk	0,010	0,16

TEST *F* UNIVARIANTES

Variable	*Suma de cuadrados entre grupos*	*Suma de cuadrados intra grupos*	*Grados de libertad*	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*F Estadístico*	*Significación*
X_9	21,682	2361,764	2 y 94	10,841	25,125	0,431	0,651
X_{10}	0,170	32,435	2 y 94	0,085	0,345	0,247	0,782

CONTRASTES DE REDUCCIÓN DE ROY BARGMAN

	Media al cuadrado	*Media al cuadrado*	F	*Grados de libertad*		*Significación*
Variable	*entre grupos*	*intra grupos*	*reducción*	*entre grupos*	*intra grupos*	*de la reducción F*
X_9	10,841	25,125	0,431	2	94	0,651
X_{10}	0,158	0,319	0,495	2	93	0,611

obtienen algunos puntos de interés. Primero, el tipo de industria afecta al nivel de uso (significación = 0,036) pero no a la satisfacción (significación = 0,328). De esta manera, cuando se evalúen conjuntamente, se obtendrá que el conjunto de variables dependientes es no significativo.

La significación práctica de demostrar un efecto sobre el nivel de uso nos puede llevar al empleo de resultados univariantes junto con los multivariantes, de forma que se puede contrastar con la segunda variable independiente —situación de la compra (X_{14})— donde los contrastes univariantes para ambas variables dependientes apoyan su efecto multivariante. Esto también se da en el análisis de reducción, que se puede interpretar como que la situación de compra afecta no solamente al conjunto de variables dependientes sino también a estas variables dependientes de forma separada, incluso después de considerar el impacto de otras variables dependientes. Estos resultados confirman las diferencias entre los impactos de las dos variables independientes analizados anteriormente.

Una consideración gerencial de los resultados

Los investigadores de HATCO realizaron una serie de ANOVAs y MANOVAs en un intento para comprender cómo los niveles de uso (X_9) y satisfacción (X_{10}) con el HATCO variaban a través de las características de las empresas involucradas (es decir, la compra de especificación (X_{11}), tipo de situación de compra (X_{14}), y tipo de industria (X_{13}).

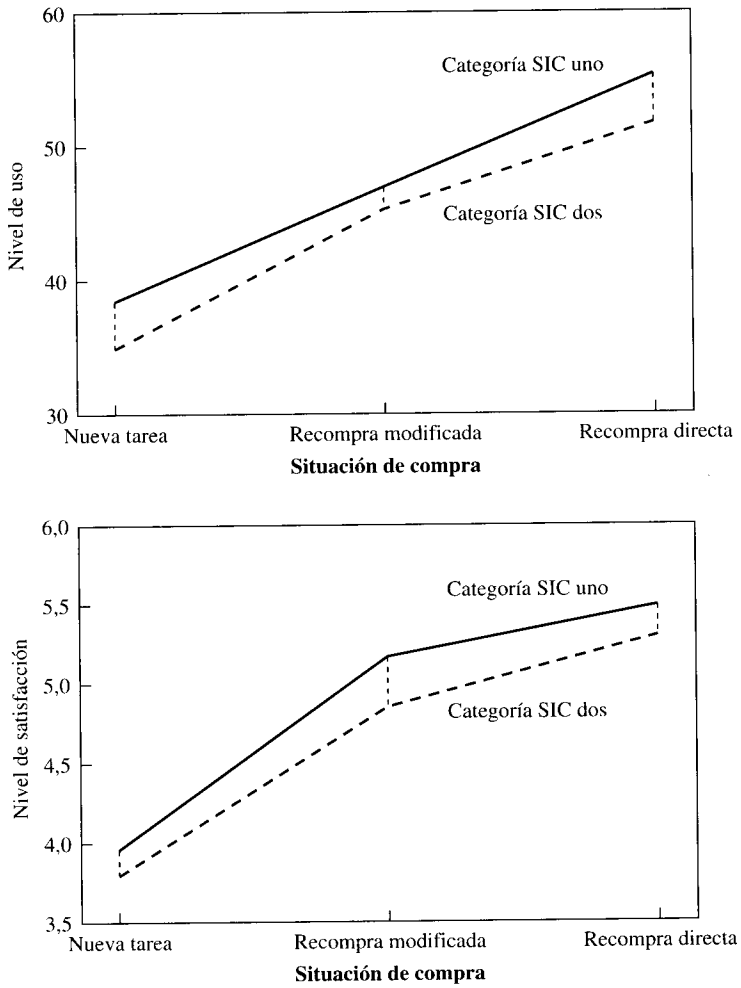


FIGURA 6.12. Dibujos de los efectos interacción en el diseño factorial MANOVA 3 x 2 para el nivel de uso y la satisfacción.

Nos centramos en los resultados multivariantes conforme se solapan con los resultados univariantes.

El primer análisis MANOVA es muy directo: ¿tiene el método de compra algún efecto sobre los niveles de satisfacción y uso? En este caso el investigador prueba si los vectores de las puntuaciones medias para cada grupo son equivalentes. Después de asegurar que se cumplen todos los supuestos, encontramos una diferencia significativa, por lo que las empresas que utilizan el análisis de valor total tienen puntuaciones de uso y satisfacción más altas cuando se comparan con las empresas que usan la compra de especificación. Junto con los resultados generales, la dirección también necesita saber si esta diferencia existe no solamente para el valor teórico sino también para las variables individuales. Los contrastes univariantes y el análisis de reducción revelan que había diferencias univariantes sustanciales tanto para los niveles de uso como satisfacción. Los resultados univariantes significativos indican a la dirección que la especificación de compra tiene un impacto único sobre los niveles de satisfacción y también de uso. Por tanto, los gerentes pueden crear estrategias únicas para cada tipo de compra de especificación para incrementar más efectivamente el uso y la satisfacción.

El próximo MANOVA sigue la misma aproximación, pero sustituye una nueva variable independiente, tipo de situación de compra, que tiene tres grupos (nueva compra, recompra modificada y recompra directa). Una vez más, la dirección se centra en los niveles de uso y satisfacción

TABLA 6.15. Tabla resumen del diseño factorial MANOVA 3 × 2: Efectos principales del tipo de industria de la situación de compra

EFECTO PRINCIPAL: TIPO INDUSTRIA (X_{13})

CONTRASTES MULTIVARIANTES DE SIGNIFICACIÓN

Nombre de contraste	*Valor*	*F Aproximada*	*Grados de libertad*		*Significación del estadístico F*
			entre grupos	*intra grupos*	
Criterio de Pillai	0,056	2,752	2	93	0,069
Traza de Hotelling	0,059	2,752	2	93	0,069
Lambda de Wilks	0,944	2,752	2	93	0,069
Mrc de Roy	0,056				

POTENCIA ESTADÍSTICA DE LOS TEST DEL MANOVA

	Efecto tamaño	*Potencia*
Criterio de Pillai	0,056	0,53
Traza de Hotelling	0,056	0,53
Lambda de Wilks	0,056	0,53

TEST *F* UNIVARIANTES

Variable	*Suma de cuadrados entre grupos*	*Suma de cuadrados intra grupos*	*Grados de libertad*	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*Estadístico F*	*Significación*
X_9	114,019	2351,764	1 y 94	114,019	25,125	4,538	0,036
X_{10}	0,872	32,435	1 y 94	0,872	0,345	2,527	0,115

CONTRASTES DE REDUCCIÓN DE ROY BARGMAN

Variable	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*Reducción F*	*Grados de libertad*		*Significación de la reducción F*
				entre grupos	*intra grupos*	
X_9	114,019	25,125	4,538	1	94	0,036
X_{10}	0,309	0,319	0,967	1	93	0,328

encontrándose de nuevo diferencias significativas. Se realizan perfiles de grupo y contraste post hoc para determinar si existe esta diferencia para los niveles de uso y satisfacción individualmente. Los contrastes post hoc son significativos, indican que existen diferencias únicas en el nivel de satisfacción, que dependen de la situación de compra, cuando se controla el nivel de uso. Desde una perspectiva gerencial, cuanto más grande es la segmentación de los clientes respecto a este tipo de variables, más adaptadas pueden ser las estrategias, dado que el director tiene un conocimiento mayor de los grupos y existen diferencias significativas entre ellas. Aunque este modelo permite al investigador considerar tres grupos, es bastante básico comparado con las numerosas situaciones que se encuentran en los entornos empresariales.

El tercer ejemplo aborda la cuestión de varias variables independientes y su impacto potencial sobre las variables dependientes. Éste es un diseño más complejo con efectos principales (el impacto sobre los niveles de satisfacción y uso) e interacciones. El uso de dos variables independientes, también permite la creación de categorías separadas con un tratamiento (grupo). En este ejemplo, las variables independientes son la situación de compra (X_{14}) y el tipo de industria (X_{13}). Las tres categorías de situación de compra se combinan con el tipo de industria para formar seis grupos. El primer paso es revisar los resultados para interacciones significativas; no se encontró ninguna. Dadas las interacciones no significativas, se puede proceder a interpretar los efectos principales. Cuando esto ocurre, el director tiene que buscar no sólo la significación estadística sino también

TABLA 6.16. Tabla resumen del diseño factorial MANOVA 3 × 2: Efectos principales de la situación de compra

EFECTO PRINCIPAL: SITUACIÓN DE COMPRA (X_{14})

CONTRASTES MULTIVARIANTES DE SIGNIFICACIÓN

Nombre de contraste	*Valor*	*F Aproximada*	*Grados de libertad*		*Significación del estadístico F*
			entre grupos	*intra grupos*	
Criterio de Pillai	0,786	30,411	4	188	0,000
Traza de Hotelling	2,852	65,601	4	184	0,000
Lambda de Wilks	0,250	45,452	4	186	0,000
Mrc de Roy	0,737				

POTENCIA ESTADÍSTICA DE LOS TEST DEL MANOVA

	Efecto tamaño	*Potencia*
Criterio de Pillai	0,393	1,00
Traza de Hotelling	0,588	1,00
Lambda de Wilks	0,500	1,00

TESTS *F* UNIVARIANTES

Variable	*Suma de cuadrados entre grupos*	*Suma de cuadrados intra grupos*	*Grados de libertad*	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*Estadístico F*	*Significación*
X_9	5580,664	2361,764	2 y 94	2790,332	25,125	111,057	0,000
X_{10}	39,250	32,435	2 y 94	19,625	0,345	56,876	0,000

CONTRASTES DE REDUCCIÓN DE ROY BARGMAN

Variable	*Media al cuadrado entre grupos*	*Media al cuadrado intra grupos*	*F reducción*	*Grados de libertad*		*Significación de la reducción F*
				entre grupos	*intra grupos*	
X_9	2790,332	25,125	111,057	2	94	0,000
X_{10}	2,793	0,319	8,752	2	93	0,000

práctica. Cuando se considera conjuntamente, ambas variables independientes tienen un impacto sobre los niveles de uso y satisfacción. Pero con relación a la significación práctica, la situación de compra tiene claramente el mayor impacto. Por tanto, el director puede determinar no solamente si una variable tiene un impacto sino también su importancia relativa con relación a las otras variables. Para la dirección de una empresa, con este método se identifican las características claves que distinguen a los clientes por lo que se refiere a su uso y evaluación de HATCO. Las estrategias que intentan tener influencia sobre uso y satisfacción deberían ser orientadas hacia la situación de compra y el tipo de industria, pero el factor más importante para tener en cuenta es la situación de compra. Por tanto, si se pueden incorporar las diferencias de mercado, sólo representarán una adición al impacto de las estrategias que son diferenciadas por una situación de compra.

Resumen

Puede resultar algo irreal suponer que la diferencia entre tratamientos experimentales se puede manifestar en una única variable dependiente considerada. Desafortunadamente, muchos investigadores tratan las situaciones de múltiple criterio como una aplicación repetida de contrastes univariantes individuales hasta analizar todas las variables dependientes. Este enfoque puede in-

crementar de forma importante los porcentajes del error de Tipo I y no tiene en cuenta la posibilidad de que alguna combinación de las variables dependientes pueda proporcionar mayor evidencia de claras diferencias entre los grupos. El empleo adecuado del MANOVA nos proporciona soluciones a ambos problemas.

Cuestiones de revisión

1. ¿Cuáles son las diferencias entre MANOVA y el análisis discriminante? ¿Qué situaciones se adecúan a cada técnica multivariante?
2. Diseñe un experimento MANOVA factorial de dos vías. ¿Cuáles son las diferentes fuentes de varianza en su experimento? ¿Qué señalaría una interacción significativa?
3. Además de la significación global o conjunta, existen al menos tres enfoques a realizar después de los contrastes: (a) usar el procedimiento de contraste de Scheffe, (b) análisis de reducción, que es similar a la regresión de reducción en donde cada sucesivo estadístico F se calcula después de eliminar los efectos de las variables dependientes previas, y (c) el examen de las funciones discriminantes. Nombre las ventajas y desventajas prácticas de cada uno de estos enfoques.
4. ¿Cómo se ve afectada la potencia estadística por las decisiones de diseño de investigación? ¿Cómo diseñaría un estudio para asegurar una potencia adecuada?
5. Describa algunas situaciones de análisis de datos de sus áreas de interés en donde el MANOVA y el MANCOVA serían apropiados. ¿Qué tipos de variables incontroladas o covariaciones pueden estar influyendo en cada una de estas situaciones?

Bibliografía

1. Anderson, T. W. (1958), *Introduction to Multivariate Statistical Analysis.* New York: Wiley.
2. Cattell, R. B., ed. (1966), *Handbook of Multivariate Experimental Psychology.* Chicago: Rand McNally.
3. Cohen, J. (1977), *Statistical Power Analysis for the Behavioral Sciences.* New York: Academic Press.
4. Cole, D. A., S. E. Maxwell, R. Avery, and E. Salas (1994), «How the Power of MANOVA Can Both Increase and Decrease as a Function of the Intercorrelations among Dependent Variables.» *Psychological Bulletin* 115: 465-74.
5. Cooley, W. W., and P. R. Lohnes (1971), *Multivariate Data Analysis.* New York: Wiley.
6. Green, P. E. (1978), *Analyzing Multivariate Data.* Hinsdale, Ill.: Holt, Rinehart, & Winston.
7. Green, P. E., and J. Douglas Carroll (1978), *Mathematical Tools for Applied Multivariate Analysis.* New York: Academic Press.
8. Green, P. E., and D. S. Tull (1979), *Research for Marketing Decisions*, 3d ed. Upper Saddle River, N.J.: Prentice Hall.
9. Hand, D. J., and C. C. Taylor (1987), *Multivariate Analysis of Variance and Repeated Measures.* London: Chapman and Hall.
10. Harris, R. J. (1975), *A Primer of Multivariate Statistics.* New York: Academic Press.
11. Hubert, C. J., and J. D. Morris (1989), «Multivariate Analysis versus Multiple Univariate Analyses.» *Psychological Bulletin* 105: 302-8.
12. Huitema, B. (1980), *The Analysis of Covariance and Alternatives.* New York: Wiley.
13. Kirk, R. E. (1982), *Experimental Design: Procedures for the Behavioral Sciences*, 2d ed. Belmont, Calif.: Brooks/Cole.
14. Koslowsky, M., and T. Caspy (1991), «Stepdown Analysis of Variance: A Refinement.» *Journal of Organizational Behavior* 12: 555-59.
15. Läuter, J. (1978), «Sample Size Requirements for the T^2 Test of MANOVA (Tables for One-Way Classification).» *Biometrical Journal* 20: 389-406.
16. Meyers, J. L. (1975), *Fundamentals of Experimental Design.* Boston: Allyn & Bacon.
17. Morrison, D. F. (1967), *Multivariate Statistical Methods.* New York: McGraw-Hill.
18. Rao, C. R. (1978), *Linear Statistical Inference and Its Application*, 2d ed. New York: Wiley.
19. Stevens, J. P. (1972), «Four Methods of Analyzing between Variations for the k-Group MANOVA Problem.» *Multivariate Behavioral Research* 7 (October): 442-54.

20. Stevens, J. P. (1980), «Power of the Multivariate Analysis of Variance Tests.» *Psychological Bulletin* 88: 728-37.
21. Tatsuoka, M. M. (1971), *Multivariate Analysis: Techniques for Education and Psychological Research.* New York: Wiley.
22. Wilks, S. S. (1932), «Certain Generalizations in the Analysis of Variance.» *Biometrika* 24: 471-94.
23. Winer, B. J. (1962), *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

LECTURA COMPLEMENTARIA

Imagen y atracción de centros comerciales suburbanos

FRANCISCO JOSÉ MÁS RUIZ
Universidad de Alicante

Investigación y Marketing, n.º 56

El objetivo fundamental de este trabajo consiste en examinar las percepciones de los consumidores acerca de los centros comerciales, así como la relación existente entre dichas percepciones y la atracción de los mismos. La hipótesis de partida es que los componentes probabilísticos originarios (distancia) y de imagen de los centros comerciales distinguiendo diversas áreas urbanas, constituyen dos elementos importantes a considerar en cualquier análisis de atracción de los mismos. La metodología aplicada utiliza el análisis discriminante para determinar las dimensiones de imagen que mejor diferencian a los consumidores habituales de los que no lo hacen asiduamente. Como resultado de la aplicación en el ámbito comercial suburbano de Alicante se detecta la existencia de diferencias estadísticamente significativas entre los centros comerciales, y que las variables que mejor discriminan la atracción varían en función de los mismos, destacando la facilidad de aparcamiento y la distancia.

1. INTRODUCCIÓN

En opinión de VÁZQUEZ (1992), es razonable pensar que los establecimientos detallistas, al igual que la mayoría de las empresas, desean proyectar una imagen concreta mediante el desarrollo de su estrategia comercial. La imagen de un establecimiento se define como el conjunto de percepciones de un consumidor sobre los atributos funcionales (bienes ofrecidos, nivel de precios, etc.,) y emocionales (ambiente, etc.,) del punto de venta. Dicha imagen es tan real como cualquier otro fenómeno psicológico y los consumidores deciden o no ir a comprar a los establecimientos detallistas en función de la imagen que tienen de los mismos. Con otras palabras, la imagen del punto de venta constituye un elemento determinante de su elección por parte del consumidor.

No obstante, una buena parte de la literatura de Marketing, defendida por las corrientes gravitacional y probabilística originariamente, ha tratado de explicar la atracción detallista obviando la imagen del punto de venta, y utilizando exclusivamente factores objetivos, como distancia al establecimiento y tamaño del mismo. Este planteamiento es criticado por DOYLE y FENWICH (1974-75) y STANLEY y SEWALL (1976), al señalar que los estudios de imagen detallista demuestran que la percepción del consumidor acerca de los comercios es un fenómeno multidimensional y que los factores de distancia y de tamaño no son siempre los mejores a la hora de predecir la atracción detallista.

Los estudios realizados sobre la imagen y la atracción de los establecimientos detallistas han tomado históricamente como unidad de referencia el punto de venta individual. Sin embargo, muy poco se ha realizado a nivel de centro comercial, entendido como conjunto de comercios localizados próximamente, cada uno de ellos de propiedad individual.

En realidad, pocos establecimientos «al detall» existen como entidades aisladas. Las sinergias derivadas de la proximidad de múltiples tiendas, el entorno legal que restringe las posibles localizaciones a los detallistas, y la disponibilidad limitada de zonas atractivas, tienden a fomentar el agrupamiento del comercio detallista en centros relativamente compactos. En esta línea, no existe un razonamiento teórico que impida extender las investigaciones de la imagen y atracción a nivel de centro comercial. De hecho, NEVIN y HOUSTON (1980) y HOWELL y ROGERS (1980) proponen que la imagen del centro comercial está significativamente relacionada con la elección del mismo, defendiendo la existencia de otras discrepancias además del tamaño y la distancia. En consecuencia, los componentes probabilísticos originarios (tamaño/distancia) y de imagen de los centros co-

merciales distinguiendo diversas áreas urbanas, constituyen dos elementos importantes a considerar en cualquier análisis de atracción de los mismos.

El propósito fundamental de este trabajo consiste en examinar las percepciones de los consumidores acerca de los centros comerciales, así como la relación existente entre dichas percepciones y la atracción de los mismos. Como ejemplo, la aplicación empírica se realiza en el contexto comercial suburbano de Alicante en 1994, teniendo en cuenta que la unidad de referencia es el centro comercial.

En lo que sigue, el resto de la investigación se ha organizado de la siguiente forma: en el apartado segundo se revisan las investigaciones empíricas realizadas sobre la influencia de la imagen en la atracción de los centros comerciales. En el tercero se expone la metodología propuesta así como su justificación. A continuación, se define la muestra, la obtención de datos, y las variables utilizadas en el contraste. La quinta sección describe los resultados obtenidos, y en la última se sintetizan las conclusiones.

2. EVIDENCIAS EMPÍRICAS PREVIAS

A pesar de la enorme importancia de la imagen para ayudar a determinar el éxito o fracaso de un centro comercial, a este elemento se le ha prestado muy poca atención en muchos estudios sobre la atracción de los mismos. Entre ellos destacan los siguientes:

Algunos investigadores han intentado clasificar el poder de las variables de imagen para explicar los niveles de captación. Una aproximación ha sido generalizar el modelo inicial de HUFF (1964) con variables de imagen y examinar así si el poder de predicción de tales modelos mejora de forma significativa.

En esta línea, GAUTSCHI (1981) estima el modelo de Huff ampliado mediante el análisis de regresión, considerando distintas variables que reflejan características de los centros comerciales y de transporte a los mismos. No obstante, únicamente investiga el efecto de cada una de las variables por separado sin tener en cuenta las posibles interrelaciones existentes entre ellas.

Una corriente complementaria defiende la ampliación de la modelización de Huff, incidiendo sobre todo en que los elementos determinantes de la imagen y selección de establecimientos detallistas implican una relación compleja entre todos los factores estudiados. Ello condiciona la metodología empleada, aplicando técnicas estadísticas de agrupamiento como el análisis factorial.

En esta línea particular, algunos autores siguen un enfoque inductivo (a través de los datos) a la hora de definir la imagen, es decir, derivan «a posteriori» la naturaleza de la imagen a partir de los resultados obtenidos de un conjunto específico de datos. En particular, especifican las dimensiones de imagen de los centros comerciales mediante la aplicación de un análisis factorial. Con posterioridad examinan la influencia de los componentes obtenidos y de la variable tamaño/distancia en un modelo de Huff ampliado. Los trabajos de NEVIN y HOUSTON (1980) y MCGOLDRICK (1992) se encuadrarían en la misma. No obstante, el estudio de McGoldrick no considera el tamaño del centro y añade ciertos rasgos personales de los individuos. Otro estudio íntimamente relacionado sería el de WEE (1986), que analiza las dimensiones de la imagen y su influencia sobre la atracción, pero no considera las variables originales del modelo de Huff, tamaño y distancia.

De otro lado, están los que toman como referencia una aproximación deductiva dirigida por la teoría al desarrollo de la imagen. En este sentido, especifican «a priori» una estructura hipotética de las dimensiones de imagen, basada en un examen teórico exhaustivo, y utilizando posteriormente un análisis factorial confirmatorio validan la estructura teórica en los centros comerciales. A continuación, agregan los factores obtenidos al modelo originario de Huff y estudian la atracción detallista a través de un análisis de regresión. El trabajo de HOWELL y ROGERS (1980) estaría en esta línea, aunque ignora las dimensiones de tamaño.

Alternativamente, HAUSER y KOPPELMAN (1979) contrastan la bondad de los análisis discriminante, factorial, y escala multidimensional, a la hora de analizar la imagen de las localizaciones comerciales. Para ello, utilizan las puntuaciones resultantes de aplicar las distintas técnicas a una serie de variables de imagen y de distancia percibidas por los consumidores sobre los centros comerciales, como dimensiones de partida en un análisis logit.

Por último, GENTRY y BURNS (1977-78) analizan las percepciones de imagen de los consumidores en relación a tres centros comerciales y utilizan el análisis discriminante para conocer las variables que mejor discriminan respecto de la frecuencia de uso de los mismos. Concluyen que la variable de distancia, medida en términos de proximidad percibida, es poco importante en la atracción del centro comercial, aunque es la que mejor discrimina.

En suma, los análisis realizados sobre la influencia de la imagen de los centros comerciales sobre la atracción ejercida por los mismos, difieren entre sí en el colectivo de entidades estudiadas, las variables seleccionadas y el sentido dado a las mismas, en las técnicas estadísticas aplicadas, y como resultado de todo ello en las conclusiones obtenidas. Por tanto, es difícil de momento extraer consecuencias acerca de la influencia de la imagen en la atracción de los centros.

Todos ellos tienen como objetivo detectar las características de imagen y de distancia que influencian la atracción de las empresas minoristas. En este artículo se propone la inclusión de la imagen detallista como forma alternativa de modelizar la atracción de dichos centros, utilizando el análisis discriminante para determinar las variables que mejor diferencian a los individuos que compran frecuentemente de los que lo hacen con menor cadencia.

3. METODOLOGÍA

La metodología de investigación desarrollada en orden a alcanzar los objetivos planteados cubre las siguientes etapas: examinar las percepciones de los consumidores sobre los centros comerciales, y analizar la relación existente entre dichas percepciones y la atracción de los mismos.

La primera etapa de la metodología propuesta, examen de las percepciones de los consumidores sobre los centros comerciales, se divide a su vez en las dos fases siguientes:

a) Descripción de las percepciones de los centros comerciales por parte de los consumidores. Para ello, se analiza el perfil de los distintos centros en términos de los valores que toma la media aritmética simple de las variables definidoras de la imagen así como de distancia para los mismos. En suma, este apartado trata de identificar la estrategia comercial que distingue a un centro de otro.

Para ello, en primer lugar, se calculan los valores medios de las variables para cada centro. En segundo lugar, se obtienen para el conjunto de las áreas los valores medios de los indicadores. Finalmente, se comparan los ratios de cada centro con el promedio de la muestra de manera que, se afirma que un centro comercial se especializa en una determinada dimensión, cuando el valor del ratio que la especifica es un 5 por ciento superior al coeficiente medio de la totalidad del equipamiento comercial de la ciudad. Con el valor un 5 por ciento inferior se puede afirmar que el centro intenta evitar la dimensión en cuestión, lo que se podría interpretar como que se requiere una experiencia de la que las empresas del centro podrían carecer.

b) El análisis descriptivo anterior se complementa con un análisis de la varianza (ANOVA) y con un análisis multivariante de la varianza (MANOVA) sobre los valores promedios de las variables en cada centro comercial. Ambos tratan de demostrar la existencia de diferencias significativas entre los distintos centros mediante el examen de la variabilidad entre los mismos, sin embargo, mientras el primero analiza cada una de las variables por separado, el segundo considera todas éstas simultáneamente.

Finalmente, la segunda fase analiza la relación existente entre imagen y atracción de los centros comerciales. Para ello, se aplica un análisis discriminante a un conjunto de variables características de la imagen de los distintos centros con el objetivo de determinar las variables que mejor diferencian a los individuos que compran frecuentemente de los que lo hacen con menor cadencia.

4. MUESTRA, OBTENCIÓN DE DATOS Y VARIABLES

El proceso metodológico presentado en el epígrafe anterior se desarrolla a continuación para el caso particular de los consumidores de cuatro centros comerciales de la ciudad de Alicante con establecimientos detallistas de productos de consumo ocasional (ropa, muebles y electrodomésticos), ejemplo atractivo para analizar las percepciones de los compradores y su relación con la atracción de los centros.

Los cuatro centros considerados distinguen las zonas céntricas de la ciudad de las ubicadas en las afueras de la misma. En particular, el «Corazón de Alicante» es un distrito central revitalizado, constituido por un gran número de tiendas especializadas, mayoritariamente asociadas entre sí para defender sus intereses. De otro lado, «Maisonnave-Oscar Esplá» está ubicado en otro céntrico núcleo urbano, aunque conformada por dos grandes almacenes, y un gran número de tiendas especializadas. Por último, las «Carreteras de San Vicente» y de «San Juan» constituyen dos centros de atracción perfectamente comunicados por autovía, que incluyen cada uno a un hipermercado y varias grandes superficies de productos ocasionales.

La información básica se obtiene de una encuesta personal con cuestionario estructurado, dirigida a una muestra de 177 individuos que participan activamente en las decisiones de compra familiares en los hogares de las ciudades de Alicante y de San Vicente del Raspeig durante el mes de noviembre de 1994. El muestreo es aleatorio polietápico y la selección de los entrevistados se realiza a través del procedimiento de rutas aleatorias. La inspección detallada de la información obtenida permitió detectar diversos errores en determinadas entrevistas, quedando eliminadas del análisis, por lo que el número de encuestas válidas se redujo considerablemente.

La variable dependiente del modelo discriminante aplicado a los diferentes centros comerciales constituye una medida de la conducta pasada de compra en cada centro. En concreto, refleja el número de veces que el entrevistado ha ido a comprar, durante los últimos doce meses, productos de consumo ocasional a cada uno de los centros. Estas frecuencias absolutas se convierten en relativas a partir del número total de viajes de compra de cada individuo en dicho año (GAUTSCHI, 1981). Se considera «viaje de compra» al realizado para buscar productos ocasionales, con independencia de que se compre o no.

Las variables independientes consideradas son las siguientes:

1. Distancia Percibida. Viene medida en forma de tiempo promedio de desplazamiento percibido por el entrevistado desde su lugar de residencia al centro comercial para diferentes medios de transporte.
2. Las variables utilizadas para definir la imagen de los centros comerciales son las siguientes: variedad de productos, profesionalidad de los vendedores, informalidad en la forma de vestir del comprador, tranquilidad en el proceso de compra, limpieza, facilidad de comunicación entre establecimientos, facilidad de aparcamiento, precios bajos en relación a la calidad ofrecida, y horarios comerciales por las tardes y fines de semana. Estas variables han sido propuestas por GAUTSCHI (1981), y se especifican a través de una serie de afirmaciones sobre las que el individuo se identifica usando una escala de cinco items, donde el cinco refleja si está totalmente de acuerdo y uno en caso contrario.

5. RESULTADOS OBTENIDOS

5.1. Análisis perceptual de los centros comerciales

El Cuadro 6.L.1. y la Figura 6.L.1 recogen los valores medios de las diferentes variables para cada uno de los centros

comerciales. En cuanto a la distancia, los resultados obtenidos muestran claramente la proximidad o lejanía de los centros respecto del núcleo urbano. En este sentido, los consumidores perciben muy próximas de su lugar de residencia a las áreas céntricas «Corazón de Alicante» y «Maisonnave-Oscar Esplá», y muy alejadas a las situadas en los extrarradios como la «Ctra. de San Vicente» y la «Ctra. de San Juan».

En relación a las restantes características de imagen, se observa que el perfil de los centros de las carreteras de «San Vicente» y de «San Juan» son muy parecidos entre si para cualquiera de las variables analizadas, pero distintos respecto de las dos áreas céntricas. Por contra, «Maisonnave Oscar Esplá» y «Corazón de Alicante» mantienen ciertas diferencias, aunque se asemejan sobre todo en las variables de informalidad en la forma de vestir, profesionalidad de los vendedores, comunicación entre establecimientos, precios y facilidad de aparcamiento.

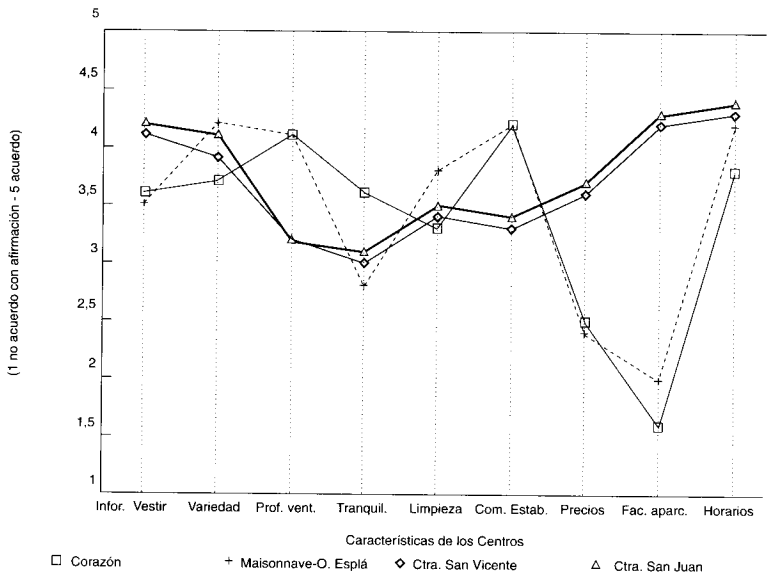
Este apartado también trata de identificar la estrategia que distingue a un Centro de otro, por lo que se comparan los valores medios de las variables de cada centro con el promedio del sector comercial en la ciudad de Alicante. Con otras palabras, resulta interesante examinar el comportamiento que presentan los centros, al estudiar los elementos potencialmente definidores de la estrategia adoptada por cada zona comercial. El Cuadro 6.L.2 muestra los diferentes modelos de competencia que utilizan los cuatro centros comerciales.

Corazón de Alicante: Este centro comercial está compuesto por tiendas tradicionales de la zona centro de la ciudad. Se trata de entidades cuyo perfil queda definido por una

CUADRO 6.L.1. Valores medios y desviaciones típicas de las variables perceptuales

Centro Comercial	*Dista.*	*Infor. Vestir*	*Varie. prtos.*	*Profe. Vende.*	*Tranq. compra*	*Limp.*	*Comun. Estab.*	*Precio*	*Fac. Aparc.*	*Hora*
Corazón	18,6	3,6	3,7	4,1	3,6	3,3	4,2	2,5	1,6	3,8
Alicante	(9,6)	(1,6)	(1,4)	(1,3)	(1,4)	(1,4)	(1,2)	(1,3)	(1,2)	(1,8)
Maisonnave	17,9	3,5	4,2	4,1	2,8	3,8	4,2	2,4	2,0	4,2
O. Esplá	(10,6)	(1,6)	(1,2)	(1,2)	(1,4)	(1,3)	(1,2)	(1,3)	(1,4)	(1,4)
Cta. San	22,9	4,1	3,9	3,2	3,0	3,4	3,3	3,6	4,2	4,3
Vicente	(11,0)	(1,4)	(1,4)	(1,4)	(1,5)	(1,4)	(1,6)	(1,3)	(1,4)	(1,4)
Cta. San	27,0	4,2	4,1	3,2	3,1	3,5	3,4	3,7	4,3	4,4
Juan	(12,0)	(1,4)	(1,3)	(1,4)	(1,5)	(1,4)	(1,6)	(1,2)	(1,3)	(1,2)
Total	**21,5**	**3,9**	**4,0**	**3,7**	**3,1**	**3,5**	**3,8**	**3,1**	**3,0**	**4,2**
	(11,4)	**(1,5)**	**(1,3)**	**(1,4)**	**(1,5)**	**(1,4)**	**(1,5)**	**(1,4)**	**(1,8)**	**(1,4)**

Fuente: Elaboración propia.



Fuente: Elaboración propia.

FIGURA 6.L.1. Comparación de centros comerciales.

estrategia que persigue unos niveles superiores de profesionalidad en sus vendedores, de tranquilidad y ausencia de multitud a la hora de realizar la compra, así como de comunicación entre los establecimientos del mismo que facilita ir de tienda en tienda. Sin embargo, entre sus inconvenientes destacan la poca variedad de productos, la menor importancia relativa de la limpieza, la escasa facilidad para aparcar, así como la dificultad para ir de compras por las tardes y fines de semana. Aunque en menor medida, también incide en dicho centro la proximidad de la distancia al consumidor, los precios elevados, y la formalidad en la manera de vestir.

Maisonnave-O. Esplá: Tiene como «núcleo» básico a dos importantes grandes almacenes españoles, que ejercen una importante atracción en el área comercial. Destaca por su modelo potenciador de la proximidad al consumidor, su variedad de productos, profesionalidad de los vendedores, facilidad de comunicación entre comercios, limpieza y distinción al estimular la formalidad en la manera de vestir. Por contra, se distingue por la reducida tranquilidad existente en la compra y por sus elevados precios para la calidad ofrecida. Aunque en menor medida, también incide la escasa facilidad de aparcamiento.

Ctra. de San Vicente: Posee como «núcleo» fundamental a un hipermercado, además de diversas grandes superficies del mueble. Quedan definidos, sobre todo, por su inferioridad en cuanto a la profesionalidad de sus vendedores y las dificultades de comunicación entre establecimientos. En menor medida, también destaca por su ubicación alejada del núcleo urbano, por los niveles reducidos de precios, así como por la facilidad de aparcamiento y de vestir informalmente.

Ctra. de San Juan: Está conformado por un hipermercado y otras grandes superficies especializadas en el equipamiento del hogar. Posee una distancia superior al promedio, y concentra su imagen en diversas variables, potenciando los bajos precios en relación con la calidad ofrecida, la informalidad en la forma de vestir, la facilidad de aparcamiento y su amplitud de horarios al abrir por las tardes y fines de semana. No obstante, también destaca por la escasa profesionalidad de sus vendedores y por la mala comunicación entre establecimientos.

Asimismo, el objetivo de esta etapa consiste en demostrar la existencia y la significatividad estadística de las diferencias en las percepciones de los consumidores sobre los centros comerciales, en términos de las dimensiones de imagen. La demostración de ambas hipótesis, por medio de un análisis de la varianza (ANOVA) y de un análisis multivariante de la varianza (MANOVA), incluye la consideración de distintas variables para reflejar la naturaleza multidimensional de la imagen empresarial.

Los resultados obtenidos permiten concluir que los centros comerciales son significativamente distintos (véase el Cuadro 6.L.3 a un nivel del 1 por ciento, dados los valores ofrecidos por los tests F para la mayoría de variables consideradas aislada y simultáneamente, en un análisis de la varianza aplicado a los valores medios de las variables de imagen en cada centro.

En concreto, el análisis multivariante de la varianza (MANOVA) se utiliza para probar si los niveles promedios de cada una de las dimensiones de imagen, difieren entre los centros comerciales. Se observan claramente diferencias significativas entre los mismos a un nivel de confianza del 99 por ciento. Consecuentemente, los resultados de este análisis son indicativos de que las dimensiones revelan diferencias entre los centros.

Asimismo, el análisis de varianza unidireccional (ANOVA) puede ayudar a examinar más detenidamente estos resultados para cada una de las medidas de forma separada. Del Cuadro 6.L.3 se desprende que existen diferencias significativas para la mayoría de dimensiones analizadas. Por contra, las variables limpieza y variedad de productos defienden débilmente la diferencia entre las percepciones a unos niveles de a inferiores al 10 por ciento. Las posibles explicaciones del comportamiento de estas últimas variables pueden venir dadas por posibles diferencias de imagen entre los establecimientos individuales de un mismo centro, como consecuencia de las distintas estrategias comerciales.

En resumen, los resultados del MANOVA y ANOVA apoyan la existencia de diferencias estadísticamente significativas entre los centros comerciales a un nivel inferior al 1 por ciento para todas las variables de forma conjunta, así como para cualquiera de ellas individualmente salvo en las dimensiones de horarios comerciales ($a < 5$ por ciento) variedad de productos y limpieza ($a < 10$ por ciento).

CUADRO 6.L.2. Modelos de competencia de los cuatro centros comerciales

Centro Comercial	*Dista.*	*Infor. Vestir*	*Varie. prtos.*	*Profe. Vende.*	*Tranq. compra*	*Limp.*	*Comun. Estab.*	*Precio*	*Fac. Aparc.*	*Hora*
Corazón arr./deb.*	–	–	–	+	+	–	+	–	–	–
Maisonnave arr./deb.	–	–	+	+	–	+	+	–	–	0
San Vicente arr./deb.	+	+	–	–	–	–	–	+	+	+
San Juan arr./deb.	+	+	+	–	0	0	–	+	+	+

* Promedio del centro está por arriba (+) o debajo (–) del promedio sectorial.
Fuente: Elaboración propia.

CUADRO 6.L.3. MANOVA y ANOVA de las características entre centros

Variable	*F*	*Prob.*
Distancia	15,062	0,000
Informalidad en la forma de vestir del comprador	4,511	0,004
Variedad de productos	2,509	0,058
Profesionalidad de los vendedores	17,429	0,000
Tranquilidad en el proceso de compra	5,209	0,002
Limpieza	2,322	0,075
Facilidad de comunicación entre establecimientos	13,614	0,000
Precios bajos en relación a la calidad	32,087	0,000
Facilidad de aparcamiento	115,592	0,000
Horarios comerciales por las tardes	2,995	0,031
MANOVA	14,666	0,000

ANOVA con 3 y 401 grados libertad; MANOVA con 30 y 1.151 grados de libertad.
Fuente: Elaboración propia.

5.2. Relación entre valoración y atracción de los centros

Sobre la base de dos grupos de compradores de los centros comerciales, más frecuentes e infrecuentes, se aplica un análisis discriminante múltiple en cada una de las áreas con el fin de examinar las diferencias más significativas entre dichos conjuntos de consumidores en términos de determinadas características de los centros. Los grupos de individuos se forman distinguiendo aquéllos cuya frecuencia relativa de viajes de compra a cada centro comercial es inferior al 20 por ciento de los que superan el 40 por ciento. De esta manera, se puede conocer las variables que ejercen influencia en la frecuencia de compra.

Las funciones de clasificación obtenidas para cada centro comercial se ofrecen en el Cuadro 6.L.4. Los coeficientes estandarizados de la función discriminante se utilizan cuando las variables han sido previamente estandarizadas, al permitir una aproximación a la importancia relativa de cada variable en la función discriminante.

Los resultados obtenidos muestran que los coeficientes más elevados corresponden a la variable facilidad de aparcamiento en «Corazón de Alicante», «Ctra. San Vicente» y «Maisonnave-O. Esplá» mientras que la distancia lo es, sobre todo, en «Maisonnave», «San Vicente» y «San Juan».

Centrándonos en los distintos centros, y por orden de importancia, las variables que mejor discriminan en la frecuencia de visita de compra a «Corazón de Alicante» son facilidad de aparcamiento, profesionalidad de los vendedores y variedad de productos. En «Maisonnave-O. Esplá» destacan la distancia, la facilidad de aparcamiento y la informalidad en la forma de vestir del comprador. En la «Ctra. de San Vicente» la facilidad de aparcamiento, la distancia, y la limpieza. Finalmente, en la «Ctra. de San Juan» sobresalen los precios bajos en relación a la calidad, la distancia, la informalidad en la forma de vestir del comprador y los horarios comerciales por las tardes.

Por otro lado, se manifiestan diferencias estadísticamente significativas entre las medias de las puntuaciones discriminantes en «Corazón de Alicante» (prob. = 0,0254), en «Ctra.

CUADRO 6.L.4. Coeficientes estandarizados de la función de clasificación de los centros comerciales

Variables	*Corazón Alicante*	*Maisonn. O. Esplá*	*Ctra. S. Vicent*	*Ctra. S.Juan*
Distancia	–0,187	0,590	–0,548	–0,738
Informalidad en la forma de vestir del comprador	0,031	0,485	0,012	0,670
Variedad de productos	0,494	–0,329	0,202	0,291
Profesionalidad de los vendedores	–0,566	–0,169	0,223	0,202
Tranquilidad en el proceso de compra	0,057	–0,294	0,190	0,191
Limpieza	–0,038	–0,209	–0,548	–0,062
Facilidad de comunicación entre establec.	–0,185	0,016	0,269	–0,306
Precios bajos en relación a la calidad	0,179	–0,367	–0,120	0,907
Facilidad de aparcamiento	0,707	0,532	0,659	–0,373
Horarios comerciales por las tardes	0,398	–0,094	0,261	–0,452

Fuente: Elaboración propia.

de San Vicente» (prob. = 0,0266), y en la «Ctra. de San Juan» (prob. = 0,0014). Asimismo, la función discriminante clasifica correctamente al 73,5 por ciento de los consumidores en «Corazón de Alicante», al 69,2 por ciento en «Maisonnave-O. Esplá», al 80,5 por ciento en la «Ctra. de San Vicente», y al 89,1 por ciento en la «Ctra. de San Juan».

6. CONCLUSIONES

La implicación de que los componentes gravitacionales y de imagen de los centros comerciales distinguiendo diversas áreas urbanas constituyen dos elementos importantes a considerar en cualquier análisis de atracción de los mismos, ha permitido analizar los fenómenos de atracción comercial en el ámbito urbano de Alicante.

La metodología empleada utiliza el análisis discriminante para determinar las variables que mejor diferencian a los compradores frecuentes de los que lo hacen con menor cadencia.

Como resultado de la aplicación empírica llevada a cabo se detecta la existencia de diferencias estadísticamente significativas entre los centros comerciales considerados en la ciudad de Alicante para las variables de imagen y de distancia.

Por otro lado, las variables que mejor discriminan la atracción de los centros comerciales varían en función de los mismos, destacando en general la facilidad de aparcamiento y la distancia. Ello permite apoyar el uso de modelos multiatributos de imagen que complementen a las dimensiones de distancia de los modelos probabilísticos originarios para explicar la atracción de compra intraurbana.

BIBLIOGRAFÍA

DOYLE, P. E. I. FENWICH, «How Store Image Affects Shopping Habits in Grocery Chains», *Journal of Retailing,* 50, invierno 1974-75, págs. 39-52.

GAUTSCHI, D. A., «Specification of Patronage Models for Retail Center Choice», *Journal of Marketing Research,* 18, 1981, págs. 162-174.

GENTRY J. W. Y A. C. BURNS, «How "Important" Are Evaluative Criteria in Shopping Center Patronage?», *Journal of Retailing,* 53(4), invierno 1977-78, págs. 73-86.

HAUSER, J. R. Y F. S. KOPPELMAN, «Alternative Perceptual Mapping Techniques: Relative Accuracy and Usefulness», *Journal of Marketing Research,* 16,1979, págs. 495- 506.

HOWELL, R. D. Y J. D. ROGERS, «Research Into Shopping Mall Choice Behavior», en K. B. Monroe (ed.), *Advances in Consumer Research,* vol. 8, Association for Consumer Research, Ann Arbor, 1980, págs. 671-676.

HUFF, D. L., «Defining and Estimating a Trading Area», *Journal of Marketing,* 28, julio 1964, págs. 34-38.

MCGOLDRICK, P., «El Desarrollo en Europa de Centros de Compra a Gran Escala Fuera de las Ciudades», en *La Distribución Comercial en el 2000,* Generalitat Valenciana, Valencia, 1992, págs. 67-96.

NEVIN, J. R. Y M. J. HOUSTON, «Image as a Component of Attraction to Intraurban Shopping Areas», *Journal of Retailing,* 56(1), 1980, págs. 77-93.

STANLEY; T. J. Y M. A. SEWALL, «Image Inputs to a Probabilistic Model: Predicting Retail Potential», *Journal of Marketing,* 40, julio 1976, págs. 48-53.

VÁZQUEZ, R., «La Imagen de la Empresa Detallista y su Percepción por los Consumidores: Una Aplicación del Análisis Factorial», *Esic-Market,* julio- septiembre 1989, págs. 49-75.

WEE, C. H., «Shopping Area Image: Its Factor Analytic Structure and Relationships with Shopping Trips and Expenditure Behavior», en R. J. Lutz (ed.), *Advances in Consumer Research,* Association for Consumer Research, Provo, 1986, págs. 49-52.

CAPÍTULO 7

Análisis conjunto

OBJETIVOS DE APRENDIZAJE

- Explicar los diferentes usos que tiene el análisis conjunto para la dirección de una empresa.
- Entender las líneas principales de la selección de variables a examinar por el análisis conjunto, así como sus valores.
- Formular el plan experimental para un análisis conjunto simple, incluyendo la creación de diseños factoriales y la comprensión del impacto de la elección a partir de un abanico de posibilidades o partiendo de la calificación como medida de preferencia.
- Evaluar la importancia relativa de las variables predictoras y de sus niveles al verse afectadas por los juicios del consumidor.
- Aplicar un simulador de elecciones a los resultados del análisis conjunto para predecir los juicios del consumidor acerca de las nuevas combinaciones de atributos.
- Examinar la implicación de elegir un modelo de efectos principales en lugar de un modelo que comprenda términos de interacción y demostrar las aproximaciones para establecer la validez de un modelo frente al otro.
- Reconocer las limitaciones del análisis conjunto tradicional y seleccionar la metodología alternativa apropiada (es decir, basada en elección o adaptativa) cuando sea necesario.

Desde la primera mitad de la década de los 70, el análisis conjunto ha suscitado una atención considerable como método para representar las decisiones de los consumidores de forma realista a través de equilibrios entre productos o servicios con muchos atributos [23]. El análisis conjunto ganó amplia aceptación y uso en muchas industrias, aumentando su utilización hasta diez veces en la década de los años 80 [66]. Durante los años 90, la aplicación del análisis conjunto aumentó aún más, extendiéndose a múltiples campos de estudio. La amplia utilización en el marketing de este tipo de análisis en el desarrollo de nuevos productos para los consumidores llevó a su

adopción en otras áreas, como el marketing industrial [41]. Este aumento de utilización en Estados Unidos ha sido paralelo con el de otras partes del mundo, particularmente en Europa [70].

Coincidiendo con este continuo crecimiento se desarrollaron métodos alternativos de construcción de las elecciones para los consumidores y de estimación de los modelos conjuntos. Muchas de las técnicas multivariantes que se discuten en este texto se establecieron en el campo de la estadística; el análisis conjunto, por otro lado, ha continuado y continuará desarrollándose en términos de su diseño, estimación y aplicaciones dentro de muchas áreas de desarrollo [11].

El uso acelerado del análisis conjunto ha coincidido con la amplia introducción de programas informáticos que integran todo del proceso, desde la generación de combinaciones de variables independientes para ser evaluadas a la creación de simuladores de elección para predecir la elección de los consumidores a lo largo de un amplio abanico de formulaciones alternativas de productos y servicios. Con un computador personal, hoy en día cualquier investigador puede acceder a varios paquetes informáticos ampliamente empleados [6, 7, 8, 28, 50, 51, 52, 53, 56, 57]. Es más, la conversión de desarrollos de investigación teórica en programas accesibles para un PC sigue produciéndose [11], y el interés en esos programas está aumentando [10, 45, 46].

El análisis conjunto está muy relacionado con la experimentación tradicional. Por ejemplo, un químico que trabaja en una planta de producción de jabón quiere conocer los efectos de la temperatura y la presión en los tanques de fabricación del jabón sobre la densidad de la pastilla de jabón resultante. Una vez que se han llevado a cabo los experimentos, se podrían analizar con los procedimientos ANOVA (análisis de varianza) tal y como se discuten en el Capítulo 6. En las situaciones que comprenden el comportamiento humano, a menudo necesitamos también «experimentos» de conducta con los factores que controlamos. Por ejemplo, ¿debería estar más o menos perfumada la pastilla de jabón? ¿Debería promocionarse como un cosmético o como un producto de higiene y/o desodorante? ¿Cuál de los tres precios propuestos debería cobrarse? Las técnicas de análisis conjunto se desarrollaron a partir de la necesidad de analizar los efectos de los factores que controlamos (variables independientes) y que a menudo están especificados cualitativamente o medidos sin fiabilidad [19, 21]. El análisis conjunto es, en realidad, una familia de técnicas y métodos, en teoría basados en los modelos de información, integración y medida funcional [38].

En términos de los modelos de dependencia básica discutidos en el Capítulo 1, el análisis conjunto puede expresarse como:

$$\underset{\text{(métrica o no métrica)}}{Y_1} = \underset{\text{(no métricas)}}{X_1 + X_2 + X_3 + \ldots + X_N}$$

El análisis conjunto se ajusta mejor para la comprensión de las reacciones de los consumidores y de las evaluaciones de combinaciones de atributos predeterminadas que representan productos o servicios potenciales. A la vez que mantiene un alto grado de realismo, ofrece al investigador una visión de la composición de las preferencias del consumidor. La flexibilidad y la unicidad del análisis conjunto surgen principalmente de: (1) su capacidad para utilizar tanto variables dependientes métricas como no métricas, (2) el uso de variables predictoras categóricas y (3) los supuestos bastante generales acerca de las relaciones entre las variables dependientes y las independientes.

¿Qué es el análisis conjunto?

El análisis conjunto es una técnica multivariante que se utiliza específicamente para entender cómo los encuestados desarrollan preferencias acerca de productos o servicios. Se basa en la simple premisa de que los consumidores evalúan el valor de un producto/servicio/idea (real o hipotética) combinando cantidades separadas de valor que proporciona cada atributo. La **utilidad**, que es la base conceptual para medir el valor en el análisis conjunto, es un juicio subjetivo de preferencia única

para cada individuo. Abarca todas las características de un producto o servicio, tanto tangible como intangible, y como tal es la medida de la preferencia global. En el análisis conjunto, la utilidad se supone que está basada en el valor asignado a cada uno de los niveles de los atributos y expresado en una relación que refleja la manera en que se formula la utilidad para cualquier combinación de atributos. Por ejemplo, podemos sumar los valores de la utilidad asociados con cada característica del producto o servicio para llegar a la utilidad conjunta. A continuación suponemos que los productos o servicios con mayores valores de utilidad son más preferidos y tienen mayor posibilidad de ser elegidos.

El análisis conjunto es el único entre los métodos multivariantes en que el investigador construye primero un conjunto real o hipotético de bienes y servicios combinando niveles escogidos de cada atributo. Estas combinaciones se presentan a los encuestados, que ofrecen sólo sus evaluaciones globales. Por tanto, el investigador pide al encuestado que realice una operación muy real —elegir entre un conjunto de productos. Los encuestados no tienen que decir nada más al investigador, como la importancia de un atributo concreto para ellos o lo bien que se ajusta el producto a un atributo específico. Dado que el investigador construye los productos o servicios hipotéticos en una forma específica, la influencia de cada atributo y de cada valor de ese atributo sobre el juicio de utilidad de un encuestado, puede ser determinante para la clasificación final del encuestado.

Para tener éxito, el investigador debe ser capaz de describir el producto o servicio en términos tanto de sus atributos como de todos los valores relevantes para cada atributo. Utilizamos el término **factor** cuando describimos un atributo específico u otra característica del producto o servicio. Los valores posibles para cada factor se denominan **niveles**. En términos conjuntos, describimos un producto o servicio respecto a *su nivel en el conjunto de factores* que lo caracteriza. Por ejemplo, el nombre de la marca y el precio pueden ser dos factores del análisis conjunto. El nombre de la marca puede tener dos niveles (marca X y marca Y), mientras que el precio puede tener cuatro niveles (39 centavos, 49 centavos, 59 centavos y 69 centavos). Cuando el investigador selecciona los factores y los niveles para describir un producto o servicio de acuerdo con un plan específico, la combinación se conoce como **tratamiento** o **estímulo**. Por tanto, un estímulo en nuestro ejemplo podría ser marca X a 49 centavos.

Un ejemplo hipotético de análisis conjunto

Como ilustración del análisis conjunto, supongamos que HATCO está intentando desarrollar una nueva máquina limpiadora industrial. Después de discutir con los representantes de ventas y grupos interesados, la ejecutiva decide que hay tres atributos importantes: ingredientes de limpieza, conveniencia del uso y nombre de la marca. Para hacer operativos estos atributos, los investigadores crean tres factores con dos niveles cada uno:

Factor	*Nivel*	
Ingredientes	Sin fosfato	Con fosfato
Forma	Líquido	Polvo
Nombre de marca	HATCO	Marca genérica

Se puede construir un hipotético producto de limpieza seleccionando un nivel de cada atributo. Para los tres atributos (factores) con dos valores (niveles), se pueden formar ocho combinaciones ($2 \times 2 \times 2$). Estos son tres ejemplos de las ocho combinaciones (estímulos):

- HATCO polvo sin fosfato
- Genérico con fosfato líquido
- Genérico sin fosfato líquido

Se pregunta a los clientes de HATCO que ordenen ocho estímulos en términos de preferencia o que los clasifiquen en una escala de preferencia (puede ser una escala de 1 a 10). Podemos ver por qué el análisis conjunto también se denomina análisis «trade-off», dado que al hacer un juicio sobre un producto hipotético, los encuestados deben considerar tanto las características «buenas» como «malas» del producto al formar una preferencia. Por tanto, los encuestados deben ponderar todos los atributos simultáneamente al hacer sus juicios.

Al construir combinaciones específicas (estímulos), el investigador está intentando entender una **estructura de preferencias** del encuestado. La estructura de preferencias «explica» no sólo lo importante que es cada factor en la decisión global, sino también cómo los diferentes niveles de un factor influyen en la formación de una preferencia conjunta (utilidad). En nuestro ejemplo, el análisis conjunto evaluaría el impacto relativo de cada nombre de marca (HATCO frente a un genérico), cada formato (polvo o líquido), y los diferentes ingredientes limpiadores (sin fosfato frente a con fosfato) en la determinación de la utilidad de una persona. Se puede pensar que esta utilidad, que representa un «valor» total o conjunto de una preferencia por un objeto, está basada en los **componentes parciales de la utilidad total** para cada nivel. La forma general de un modelo conjunto puede verse así:

$$\begin{aligned}(\text{Valor total del producto})_{ij\ldots n} = {} & \text{Componente parcial de utilidad total de nivel } i \text{ para factor 1} \\ & + \text{Componente parcial de utilidad total de nivel } j \text{ para factor 2+} \ldots \\ & + \text{Componente parcial de utilidad total de nivel } n \text{ para factor } m\end{aligned}$$

donde el producto o servicio tiene m atributos, y cada atributo n niveles. El producto consiste en el nivel i del factor 2, el nivel j del factor 2 y así sucesivamente, hasta el nivel n del factor m.

En nuestro ejemplo, un modelo aditivo simple representaría la estructura de preferencia del limpiador industrial basándose en tres factores (utilidad = efecto de marca + efecto del ingrediente + efecto de forma). La preferencia por un producto limpiador específico puede calcularse directamente a partir de los valores de los componentes parciales de la utilidad total. Por ejemplo, la preferencia de HATCO del polvo libre de fosfato es:

$$\begin{aligned}\text{Utilidad} = {} & \text{Componente parcial de utilidad total de marca HATCO} \\ & + \text{Componente parcial de utilidad total de ingredientes limpiadores sin fosfatos} \\ & + \text{Componente parcial de utilidad total del polvo}\end{aligned}$$

Con las estimaciones de los componentes parciales de la utilidad total, la preferencia de un individuo puede estimarse para cualquier combinación de factores. Es más, la estructura de preferencia revelaría qué factores son más importantes en la determinación de la utilidad conjunta y la elección del producto. Las elecciones de varios encuestados podrían combinarse también para representar el entorno competitivo a que se enfrenta el producto en el «mundo real».

Un ejemplo empírico

Para ilustrar un análisis conjunto sencillo, supongamos que el experimento del limpiador industrial se dirige a encuestados que compran productos industriales. Se presentan a cada encuestado 8 descripciones de productos limpiadores (estímulos) y se les pide que los clasifiquen en orden de preferencia de compra (1 = más preferido y 8 = menos preferido). Los ocho estímulos se describen en la Tabla 7.1, junto con las clasificaciones de las órdenes dadas por dos encuestados.

En la medida en que examinamos las respuestas para el encuestado 1, vemos que las calificaciones para los estímulos con ingredientes sin fosfato son las más altas posibles (1, 2, 3 y 4), mientras que los ingredientes con fosfato tienen las cuatro calificaciones más bajas (5, 6, 7 y 8). Por tanto, el ingrediente libre de fosfato es más preferido al limpiador con fosfato. Esto puede contrastarse con las calificaciones de las dos marcas, que muestran una mezcla de valores altos y bajos para cada marca. Suponiendo que se aplica el modelo básico (un modelo aditivo), podemos calcular el impacto de cada nivel como diferencias (desviaciones) de la calificación media global (los lectores pueden notar que esto es análogo a una regresión múltiple con variables dummy o ANOVA). Por

TABLA 7.1. Descripciones de estímulos y calificaciones de los encuestados para el análisis conjunto del limpiador industrial

	Descripciones de estímulos			*Calificaciones de los encuestados*	
	Forma	*Ingredientes*	*Marca*	*Encuestado* 1	*Encuestado* 2
1	Liquida	Sin fosfatos	HATCO	1	1
2	Liquida	Sin fosfatos	Genérico	2	2
3	Liquida	Con fosfatos	HATCO	5	3
4	Liquido	Con fosfatos	Genérico	6	4
5	Polvo	Sin fosfatos	HATCO	3	7
6	Polvo	Sin fosfatos	Genérico	4	5
7	Polvo	Con fosfato	HATCO	7	8
8	Polvo	Con fosfato	Genérico	8	6

ejemplo, las calificaciones medias para los dos ingredientes limpiadores (sin fosfatos frente a con fosfatos) para el encuestado 1 son:

Sin fosfatos: (1 + 2 + 3 +4)/4 = 2,5
Con fosfatos: (5 + 6 +7 +8)/4 = 6,5

Siendo el rango medio del octavo estímulo de 4,5 [(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8)/8 = 36/8 = 4,5], el nivel libre de fosfatos tendría entonces una desviación de –2,0 (2,5 – 4,5) respecto de la media global, mientras que el nivel con fosfatos tendría una desviación de +2,0 (6,5 – 4,5). En la Tabla 7.2 se dan las calificaciones medias y las desviaciones para cada factor de la calificación media conjunta (4,5) para los encuestados 1 y 2. En nuestro ejemplo, utilizamos los números bajos para indicar elevadas calificaciones y un estímulo más preferido (es decir, 1 = más preferido). Cuando la preferencia medida está inversamente relacionada con la preferencia, como ocurre aquí, invertimos

Tabla 7.2. Calificaciones medias y desviaciones para los encuestados 1 y 2

Nivel de factor	*Calificaciones para los estímulos*	*Calificaciones medias del nivel*	*Desviaciones de la calificación global*[a]
Encuestado 1			
Forma			
Líquido	1, 2, 5, 6	3.5	−1,0
Polvo	3, 4, 7, 8	5.5	+1,0
Ingredientes			
Sin fosfatos	1, 2, 3, 4	2.5	−2,0
Con fosfatos	5, 6, 7, 8	6.5	+2,0
Marca			
HATCO	1, 3, 5, 7	4.0	−0,5
Genérica	2, 4, 6, 8	5.0	+0,5
Encuestado 2			
Forma			
Líquido	1, 2, 3, 4	2.5	−2,0
Polvo	5, 6, 7, 8	6.5	+2,0
Ingredientes			
Sin fosfato	1, 2, 5, 7	3.75	−0,75
Con fosfatos	3, 4, 6, 8	5.25	+0,75
Marca			
HATCO	1, 3, 7, 8	4.75	+0,25
Genérica	2, 4, 5, 6	4.25	−0,25

[a]Desviación calculada como: desviación = nivel de la calificación media – calificación media global (4,5). Nótese que las desviaciones negativas implican clasificaciones más preferidas.

los signos de las desviaciones en los cálculos de los componentes parciales de la utilidad total, de tal forma que las desviaciones positivas se asociarán con componentes parciales de la utilidad total que indica una mayor preferencia.

Los componentes parciales de la utilidad total de cada nivel de calculan en cuatro pasos:

Paso 1: elevar al cuadrado de las desviaciones y hallar la suma para todos los niveles.
Paso 2: calcular un valor estandarizado que sea igual al número total de niveles dividido por la suma de los cuadrados de las desviaciones.
Paso 3: se estandariza cada desviación al cuadrado multiplicándola por el valor estandarizado.
Paso 4: estimar el componente parcial de la utilidad total tomando la raíz cuadrada de la desviación al cuadrado estandarizada.

Examinemos cómo calcularíamos el componente parcial de la utilidad total del primer nivel de ingredientes (sin fosfato) del encuestado 1. Se elevan al cuadrado las desviaciones desde 2,5. Las desviaciones al cuadrado se suman (10,5). El número de niveles es seis (3 factores con 2 niveles cada uno). Por tanto, se calcula el valor estandarizado como 0,571 (6/10,5 = 0,571). La desviación al cuadrado de sin fosfatos (2^2; recordemos que se invierten los signos) se multiplica por 0,571 para llegar a 2,284 ($2^2 \times 0{,}571 = 2{,}284$). Finalmente, para calcular la componente parcial de la utilidad total para este nivel, tomamos a continuación la raíz cuadrada de 2,284 para un valor de 1,1511. Este proceso arroja los componentes parciales de la utilidad total para cada nivel para los encuestados 1 y 2, como muestra la Tabla 7.3.

Dado que las estimaciones de los componentes parciales de la utilidad total están en una escala común, podemos calcular la importancia relativa de cada factor directamente. La importancia de un factor se representa por el rango de sus niveles (es decir, la diferencia entre los valores más alto y más bajo) dividido por la suma de los rangos para todos los factores. Por ejemplo, para el encuestado 1, los rangos son 1,512 [0,756 – (–0,756)], 3,022 [1,511 – (–1,511)], y 0,756 [0,378 – (–0,378)]. La suma total de los rangos es 5,290. La importancia relativa de la forma, ingredientes y marcas se calcula como 1,512/5,290, 3,022/5,290 y 0,756/5,290, ó 28,6 por ciento, 57,1 por ciento y 14,3 por ciento, respectivamente. Podemos seguir el mismo procedimiento para el segundo encuestado y calcular la importancia de cada factor, con los resultados de forma (66,7 por ciento), ingredientes (25 por ciento) y marca (8,3 por ciento). Estos cálculos para los encuestados 1 y 2 se muestran también en la Tabla 7.3.

Para examinar la capacidad de este modelo para predecir las elecciones reales de los encuestados, predecimos el orden de preferencias sumando los componentes parciales de la utilidad total para las diferentes combinaciones de los niveles de factor y a continuación se ordenan las puntuaciones resultantes. Los cálculos para ambos encuestados del total de los ocho estímulos se muestra en la Tabla 7.4. Comparando la predicción del orden de preferencia con el orden de preferencia real del encuestado obtenemos la precisión predictiva. Nótese que los valores totales de los componentes parciales de la utilidad total no tienen un significado real excepto como un medio de desarrollar el orden de preferencias y, como tal, no se comparan para los encuestados. Los órdenes de preferencia real y prevista para ambos encuestados se dan en la Tabla 7.4.

Los componentes parciales de la utilidad total estimados predicen perfectamente el orden de preferencia para el encuestado 1. Esto indica que la estructura de preferencias se representó con éxito en las estimaciones de los componentes parciales de la utilidad total y que el encuestado realiza elecciones consistentes con la estructura de preferencias. La necesidad de consistencia se ve cuando se observan los rangos para el encuestado 2. Por ejemplo, el rango medio de la marca genérica es más bajo que el de HATCO (referido en la Tabla 7.2), lo que significa que, si todo permanece constante, los estímulos de la marca genérica son más preferidos. Sin embargo, examinando los órdenes de rangos, no ocurre siempre así. Los estímulos 1 y 2 son iguales excepto para el nombre de la marca, aunque HATCO es más preferido. También ocurre para los estímulos 3 y 4. Sin embargo, el orden correcto (genérico preferido sobre HATCO) es visto con los pares de estímulos5-6 y 7-8. Por tanto, la estructura de preferencias de los componentes parciales de la utilidad total tendrían dificultades para predecir la elección de esta pauta. Cuando comparamos los rangos de orden previstos y efectivos (véase Tabla 7.4), vemos que las elec-

TABLA 7.3. Componentes parciales de la utilidad total estimados e importancia del factor para los encuestados 1 y 2

	Componentes parciales de la utilidad				*Cálculo de la Importancia*	
Nivel de factor	*Desviación inversa* [a]	*Desviación al cuadrado*	*Desviación estandarizada* [b]	*Estimación comp.parc. ut* [c]	*Rango de com. par. ut.*	*Importancia de factor* [d]
Encuestado 1						
Forma						
Líquido	+1,0	1,0	+0,571	+0,756	1,512	28,6%
Polvo	−1,0	1,0	−0,571	−0,756		
Ingredientes						
Sin fosfatos	+2,0	4,0	+2,284	+1,511	3,022	57,1%
Con fosfatos	−2,0	4,0	−2,284	−1,511		
Marca						
HATCO	+0,5	0,25	+0,143	+0,378	0,756	14,3%
Genérica	−0,5	0,25	−0,143	−0,378		
Suma de los cuadrados de las desviaciones		10,5				
Valor estandarizado[e]		0,571				
Suma de los rangos de los componentes parciales de la utilidad total					5,290	
Encuestado 2						
Forma						
Líquido	+2,0	4,0	+2,60	+1,612	3,224	66,7%
Polvo	−2,0	4,0	−2,60	−1,612		
Ingredientes						
Sin fosfatos	+0,75	0,5625	+0,365	+0,604	1,208	25,0%
Con fosfato	−0,75	0,5625	−0,365	−0,604		
Marca						
HATCO	−0,25	0,0625	−0,04	−0,20	0,400	8,3%
Genérica	+0,25	0,0625	+0,04	+0,20		
Suma de los cuadrados de las desviaciones		9,25				
Valor estandarizado		0,649				
Suma de los rangos de los componentes parciales de la utilidad total					4,832	

[a] Las desviaciones son las inversas para indicar un alto grado de preferencia para valores bajos. El signo de la desviación se utiliza para indicar el signo de la estimación de los componentes parciales de la utilidad total.
[b] Desviaciones estandarizadas igual al valor estandarizado por la desviación al cuadrado.
[c] Componentes parciales de la utilidad total estimados igual a la raíz cuadrada de la desviación estandarizada.
[d] Importancia del factor igual al rango de un factor dividido por la suma de los rangos para todos los factores, multiplicados por 100 para ofrecer un porcentaje.
[e] Valor estandarizado igual al número de niveles (2 + 2 + 2 = 6) dividido por la suma de las desviaciones al cuadrado.

ciones del encuestado 2 en ocasiones no están bien previstas, pero muy a menudo sólo por una posición debido al efecto de la marca. Por tanto, podemos concluir que la estructura de preferencia es una representación adecuada del proceso de elección para los factores más importantes, pero que no predice perfectamente la elección del encuestado 2, como lo hace para el encuestado 1.

TABLA 7.4. Componentes parciales estimados de la utilidad total y comparación de la clasificación de preferencias estimadas y reales

Descripción de estímulo			*Estimaciones de los componentes de la utilidad total*				*Clasificación de preferencias*	
Tamaño	*Ingredientes*	*Marca*	*Tamaño*	*Ingredientes*	*Marca*	*Total*	*Estimada*	*Real*
Encuestado 1								
Líquido	Sin fosfatos	HATCO	0,756	1,511	0,378	2,645	1	1
Líquido	Sin fosfatos	Genérica	0,756	1,511	−0,378	1,889	2	2
Líquido	Con fosfatos	HATCO	0,756	−1,511	0,378	−0,377	5	5
Líquido	Con fosfatos	Genérica	0,756	−1,511	−0,378	−1,133	6	6
Polvo	Sin fosfatos	HATCO	−0,756	1,511	0,378	1,133	3	3
Polvo	Sin fosfatos	Genérica	−0,756	1,511	−0,378	0,377	4	4
Polvo	Con fosfatos	HATCO	−0,756	−1,511	0,378	−1,889	7	7
Polvo	Con fosfatos	Genérica	−0,756	−1,511	−0,378	−2,645	8	8
Encuestado 2								
Líquido	Sin fosfatos	HATCO	1,612	0,604	−0,20	2,016	2	1
Líquido	Sin fosfatos	Genérica	1,612	0,604	0,20	2,416	1	2
Líquido	Con fosfatos	HATCO	1,612	−0,604	−0,20	0,808	4	3
Líquido	Con fosfatos	Genérica	1,612	−0,604	0,20	1,208	3	4
Polvo	Sin fosfatos	HATCO	−1,612	0,604	−0,20	−1,208	6	7
Polvo	Sin fosfatos	Genérica	−1,612	0,604	0,20	−0,808	5	5
Polvo	Con fosfatos	HATCO	−1,612	−0,604	−0,20	−2,416	8	8
Polvo	Con fosfatos	Genérica	−1,612	−0,604	0,20	−2,016	7	6

La utilización del análisis conjunto en la práctica empresarial

Antes de discutir las bases estadísticas del análisis conjunto, deberíamos entender esta técnica en términos de su papel en la toma de decisiones y el desarrollo de una estrategia. El sencillo ejemplo que acabamos de discutir presenta algunos de los beneficios básicos del análisis conjunto. La flexibilidad del análisis conjunto da pie a su aplicación en casi cualquier área en la que se estudie el proceso de toma de decisiones. El análisis conjunto supone que cualquier conjunto de objetos (es decir, marcas, compañías) o conceptos (por ejemplo, posición, beneficios, imágenes) se evalúe como un conjunto de atributos. Una vez que se ha determinado la contribución de cada factor a la valoración global del consumidor, el investigador de marketing podría a continuación:

1. Definir el objeto o concepto con la combinación óptima de factores.
2. Mostrar las contribuciones relativas de cada atributo y cada nivel de valoración conjunta del objeto.
3. Usar estimaciones de los juicios del comprador o del cliente para predecir preferencias entre objetos con diferentes conjuntos de características (manteniendo constante el resto).
4. Formar grupos aislados de clientes potenciales que dan diferentes niveles de importancia a las características para definir segmentos potenciales altos y bajos.
5. Identificar las oportunidades de marketing explorando el mercado potencial para una combinación de características no disponible en ese momento.

El conocimiento de la estructura de preferencias para cada individuo permite al investigador una flexibilidad casi ilimitada en el examen tanto de las reacciones agregadas como individuales para un amplio rango de aspectos relacionados con los productos o servicios. Examinaremos algunas de las aplicaciones más habituales al final de este capítulo.

Comparación del análisis conjunto con otros métodos multivariantes

El análisis conjunto difiere de otras técnicas multivariantes en tres áreas distintas: (1) su naturaleza de técnica de descomposición, (2) el hecho de que las estimaciones pueden hacerse a nivel individual, y (3) su flexibilidad en términos de relaciones entre variables dependientes e independientes.

Técnicas de composición frente a las técnicas de descomposición

El análisis conjunto se denomina **modelo de descomposición** debido a que el investigador necesita sólo conocer una preferencia global del encuestado para un objeto creado por el investigador a través de la especificación de valores (niveles) para cada atributo (factor). De este modo, el análisis conjunto puede *descomponer* la preferencia para determinar el valor de cada atributo. El análisis conjunto difiere de los **modelos de composición** tales como el análisis discriminante y muchas aplicaciones de la regresión, en los que el investigador obtiene calificaciones del encuestado sobre muchas características del producto (por ejemplo, predisposición hacia un color, estilo, características específicas) y a continuación las relaciona con alguna calificación de preferencia global para desarrollar un modelo predictivo. El investigador no conoce de antemano las calificaciones sobre las características del producto, pero las obtiene del encuestado. Con la regresión y el análisis discriminante las calificaciones del encuestado y las preferencias conjuntas se analizan para «componer» la preferencia conjunta a partir de las evaluaciones del encuestado del producto en *cada* atributo.

Especificación del valor teórico en el análisis conjunto

El análisis conjunto emplea un valor teórico muy similar en forma al que hemos visto en otras técnicas multivariantes. El **valor teórico conjunto** es una combinación lineal de efectos de las variables independientes (factores) sobre una variable dependiente. La diferencia importante está en que en el valor teórico conjunto, el investigador especifica tanto las variables independientes (factores) como sus valores (niveles). La única información proporcionada por el encuestado es la medida dependiente. Los niveles especificados por el investigador se utilizan en el análisis conjunto para descomponer las respuestas del encuestado en efectos para cada nivel, de la misma forma que se hace en el análisis de la regresión para cada variable independiente. Este aspecto ilustra las características comunes compartidas por el análisis conjunto y la experimentación, donde el diseño del proyecto es un paso crítico para la realización. Por ejemplo, si una variable o efecto no se anticipó en el diseño de investigación, entonces no se encuentra disponible para el análisis. Por esta razón, un investigador puede estar tentado de incluir un número de variables que *pueden* ser relevantes. Sin embargo, el análisis conjunto está limitado en el número de variables que se pueden incluir, de tal forma que el investigador no puede incluir cuestiones adicionales para compensar una falta de concepción clara del problema.

Modelos separados para cada individuo

El análisis conjunto difiere de casi cualquier otro método multivariante en que puede llevarse a cabo a nivel individual, lo que significa que el investigador genera un «modelo» separado para predecir las preferencias de *cada* encuestado. La mayoría de los métodos multivariantes toman una úni-

ca medida de preferencia (observación) de cada encuestado y a continuación desarrollan el análisis considerando a todos los encuestados simultáneamente. De hecho, muchos métodos *requieren* que un encuestado ofrezca una única observación (el supuesto de independencia) y a continuación desarrollan un modelo común para todos los encuestados, ajustando a cada encuestado con varios grados de precisión (representada por los errores de precisión para cada observación, tal como los residuos en una regresión). En el análisis conjunto, sin embargo, las estimaciones pueden ser hechas de forma individual (desagregadas) o de grupos de individuos representando un segmento de mercado o del mercado entero (agregado). A nivel desagregado, cada encuestado califica los suficientes estímulos para que el análisis sea realizado para cada persona. La precisión predictiva se calcula para cada persona, en lugar de hacerlo sólo para la muestra total. Los resultados individuales pueden agregarse para componer un modelo conjunto.

Muchas veces, sin embargo, el investigador selecciona un método de análisis agregado que realice la estimación de los componentes parciales de la utilidad total para el grupo de encuestados en conjunto. El análisis agregado puede proporcionar: (1) un medio para reducir la tarea de recogida de datos mediante diseños más complejos (discutidos en secciones posteriores), (2) métodos para estimar interacciones (es decir, análisis conjunto basado en elecciones), y (3) mayor eficacia estadística al utilizar más observaciones en la estimación. Al seleccionar entre el análisis conjunto agregado y el desagregado, el investigador debe sopesar los beneficios obtenidos por los métodos desagregados frente a las perspectivas proporcionadas por los modelos aislados obtenidos por los métodos desagregados.

Tipos de relaciones

El análisis conjunto no está limitado a los tipos de relaciones exigidas entre las variables dependientes e independientes. Como se discutió en los capítulos anteriores de la Sección 2, la mayoría de los métodos de dependencia suponen que existe una relación lineal cuando la variable dependiente aumenta (o disminuye) en cantidades iguales para cada unidad en la variable dependiente. El análisis conjunto, sin embargo, puede hacer predicciones separadas para los efectos de cada nivel sobre la variable independiente y no supone que estén relacionadas en modo alguno. El análisis conjunto puede tratar fácilmente con relaciones no lineales —incluso la compleja relación curvilineal, en la cual un valor es positivo, el siguiente negativo y el tercero positivo otra vez. Como discutiremos después, la simplicidad y flexibilidad del análisis conjunto comparado con los otros métodos multivariantes se basa en el número de supuestos hechos por el investigador.

Diseño de un experimento de análisis conjunto

Aunque el análisis conjunto pide poco al encuestado en términos tanto del número como del tipo de respuestas, el investigador debe tomar cierto número de decisiones clave en el diseño del experimento y el análisis de los resultados. La Figura 7.1 (pasos 1-3) y la Figura 7.4 (pasos 4-7) muestran los pasos generales que se siguen en el diseño y ejecución de un experimento de análisis conjunto. La discusión sigue el paradigma de construcción de modelos introducido en el Capítulo 1. El proceso de decisión se inicia con una especificación de los objetivos del análisis conjunto. Dado que el análisis conjunto es muy parecido a un experimento, la concepción de la investigación es crítica para su éxito. Una vez que se han definido los objetivos, se tratan los temas relacionados con el diseño de la investigación en sí, y se evalúan los supuestos. El proceso de decisión considera a continuación la estimación efectiva del análisis conjunto, la interpretación de los resultados y los métodos utilizados para validar los resultados. La discusión termina con un examen del uso de los resultados del análisis conjunto en ulteriores análisis tales como la segmentación de mercado y la elección de simuladores. Cada una de estas decisiones proviene de la utilización de la cuestión a investigar y el análisis conjunto como herramienta para entender las pre-

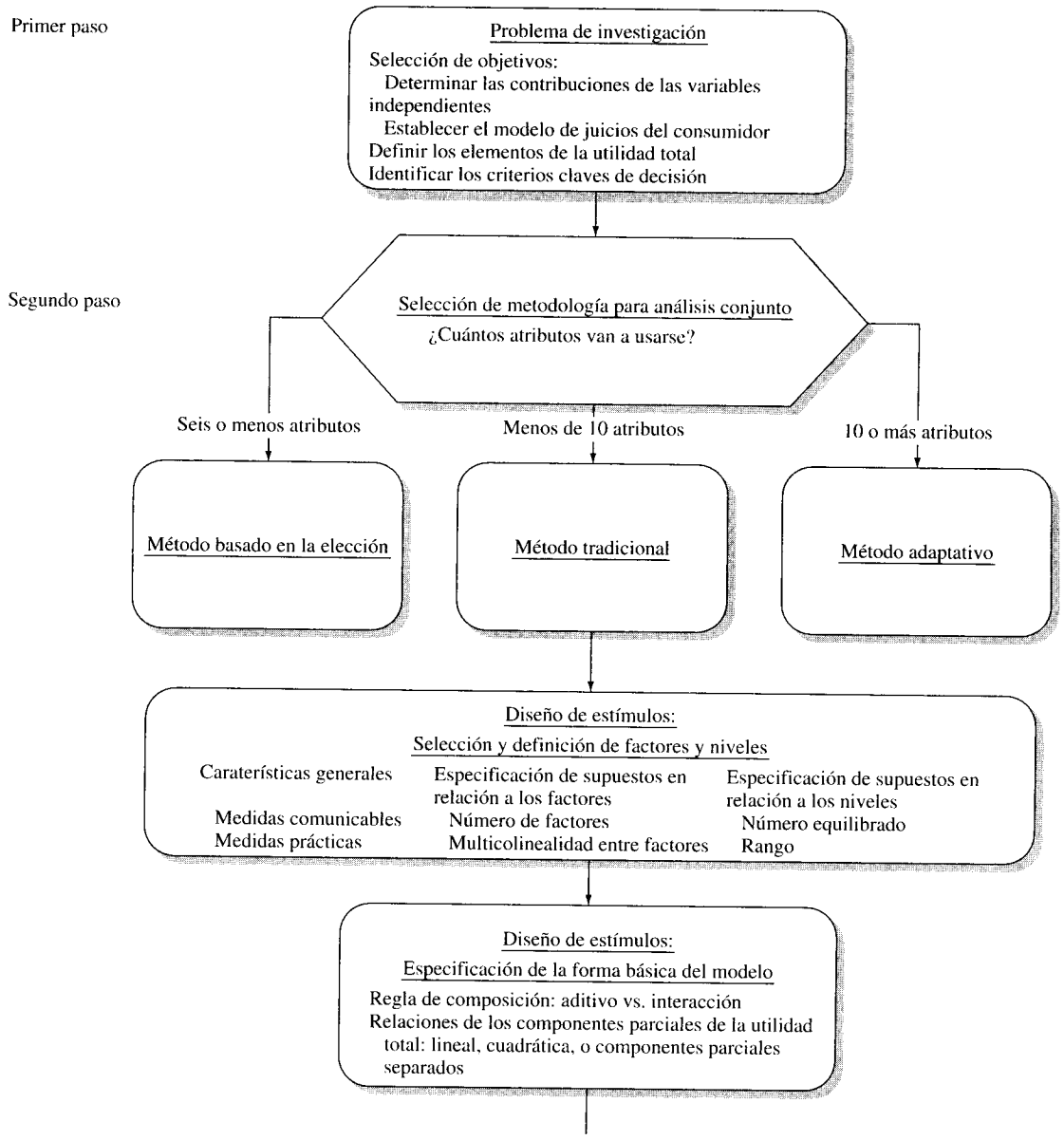
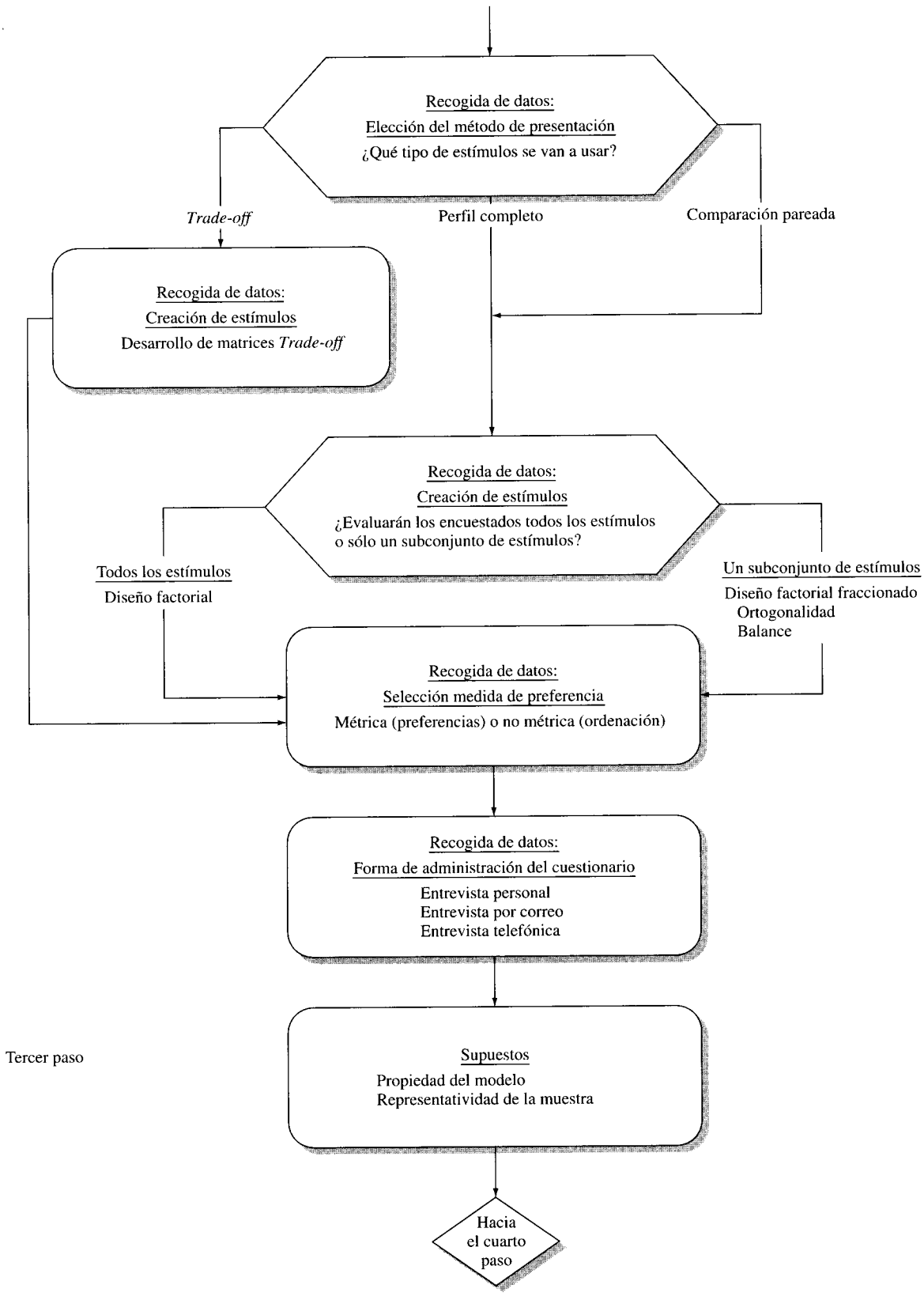


FIGURA 7.1. Pasos 1-3 del diagrama de decisión del análisis conjunto (la figura continúa en la página siguiente)

ferencias de los encuestados y su proceso de valoración. Seguimos esta discusión de la aproximación de construcción de modelos mediante el examen de dos metodologías alternativas de análisis conjunto (basado en elección y conjunto adaptativo), que se comparan a continuación con los temas que se tratan aquí del análisis conjunto tradicional.

Primer paso: Objetivos del análisis conjunto

Como ocurre con cualquier análisis estadístico, el punto de partida es la cuestión a investigar. En el análisis conjunto, el diseño experimental en las decisiones del consumidor tiene dos objetivos:


Recogida de datos:
Elección del método de presentación
¿Qué tipo de estímulos se van a usar?
Trade-off
Perfil completo
Comparación pareada
Recogida de datos:
Creación de estímulos
Desarrollo de matrices Trade-off
Recogida de datos:
Creación de estímulos
¿Evaluarán los encuestados todos los estímulos o sólo un subconjunto de estímulos?
Todos los estímulos
Diseño factorial
Un subconjunto de estímulos
Diseño factorial fraccionado
Ortogonalidad
Balance
Recogida de datos:
Selección medida de preferencia
Métrica (preferencias) o no métrica (ordenación)
Recogida de datos:
Forma de administración del cuestionario
Entrevista personal
Entrevista por correo
Entrevista telefónica
Tercer paso
Supuestos
Propiedad del modelo
Representatividad de la muestra
Hacia el cuarto paso

1. *Determinar las contribuciones de las variables predictor y sus niveles en la determinación de las preferencias del consumidor.* Por ejemplo, ¿en qué medida contribuye el precio a la disposición a comprar un producto? ¿Qué nivel de precio es el mejor? ¿En qué medida se puede tener en cuenta el cambio en la disposición a comprar jabón por la diferencia en los niveles de precios?
2. *Establecer un modelo válido de los juicios del consumidor.* Los modelos válidos nos permiten predecir la aceptación por parte del consumidor de cualquier combinación de atributos, incluso de aquellos no originariamente evaluados por los consumidores. Para hacerlo, los temas a considerar serían los siguientes: ¿indican las respuestas del encuestado una relación lineal entre las variables predictor y las elecciones? ¿Es suficiente un modelo simple que «sume» los valores de cada atributo o necesitamos añadir evaluaciones más complejas de preferencia para reflejar el proceso de valoración más adecuadamente?

Los encuestados reaccionan sólo a lo que el investigador les proporciona en términos de estímulos (combinaciones de atributos). ¿Son estos los atributos que se utilizan en realidad en el proceso de toma de decisiones? ¿Son otros atributos, particularmente atributos de una naturaleza más cualitativa tales como reacciones emotivas, también importantes? Estas y otras consideraciones exigen que la cuestión a investigar se formule alrededor de dos temas principales: (1) ¿es posible describir todos los atributos que dan utilidad o valor al producto o servicio que se está estudiando? (2) ¿cuáles son los criterios clave implicados en el proceso de elección para este tipo de producto o servicio? Estas cuestiones necesitan ser resueltas antes de proceder a la fase de diseño del análisis conjunto, dado que ofrecen una guía fundamental en cada paso.

Definición de la utilidad total del objeto

El investigador debe en primer lugar asegurarse de definir la utilidad total del objeto. Para representar el proceso de valoración del encuestado con precisión, deberían incluirse todos los atributos que potencialmente *crean* o *sustraen* utilidad al producto o servicio. Es esencial que se consideren tanto los factores positivos como los negativos, debido a que: (1) centrarse sólo en los valores positivos distorsionaría seriamente los juicios de los encuestados y (2) los encuestados pueden emplear inconscientemente factores negativos, incluso aunque no los proporcionen en la encuesta, que invalidarían el experimento. Por ejemplo, si se acude a dinámicas de grupo exploratorias para valorar los tipos de características consideradas cuando se evalúa el objeto, el investigador debe asegurarse que estudia lo que hace el objeto poco atractivo o muy atractivo. Afortunadamente, la omisión de un factor aislado sólo tiene un impacto pequeño sobre la estimación de otros factores cuando se utiliza en un modelo aditivo [49], pero puede tener algún impacto si es importante.

Especificación de los factores determinantes

Además, el investigador debe asegurarse de incluir todos los factores determinantes (deducidos del concepto de atributos determinantes [4]). El objetivo es incluir los factores que mejor *diferencian* entre los objetos. Muchos atributos pueden ser considerados importantes pero también pueden no diferenciarse en la realización de elecciones porque no varían sustancialmente entre objetos. Por ejemplo, la seguridad en el automóvil es un atributo muy importante, pero no sería determinante en la mayoría de los casos porque todos los automóviles cumplen estrictamente los reglamentos administrativos y por tanto se consideran seguros, al menos en un nivel aceptable. Sin embargo, otras características, como consumo de gasolina, rendimiento o precio son más importantes *y* probablemente serán mucho más utilizados a la hora de elegir entre diferentes modelos de coches. Por tanto, el investigador debería siempre intentar identificar las variables determinantes claves porque son las piedras angulares en la decisión.

Segundo paso: Diseño del análisis conjunto

Una vez resueltos los problemas que se desprenden de los objetivos de investigación, el investigador desplaza su atención a los temas particulares comprendidos en el diseño y ejecución del experimento de análisis conjunto. En primer lugar, ¿cuál de los diversos métodos conjuntos alternativos debe elegirse? Con el tipo de modelo seleccionado, existen temas específicos del diseño que hay que resolver. Por ejemplo, ¿cómo decide uno qué combinaciones específicas de niveles de atributos se presentan al encuestado para que las evalúe? Además de especificar las combinaciones (estímulos), el investigador debe decidir también sobre temas tales como qué atributos se incluyen, cuántos niveles de cada uno, cómo medir la preferencia y la recogida de datos y qué procedimiento de estimación se utiliza. El diseño de estos aspectos es quizá la fase más importante del análisis conjunto, de tal forma que los problemas que genera un estudio mal diseñado no pueden arreglarse después en su administración. Por tanto, el investigador debe prestar una atención particular a los temas relacionados con la construcción y administración del experimento realizado con análisis conjunto.

Selección de una metodología para el análisis conjunto

Después de que el investigador ha determinado los atributos básicos que constituyen la utilidad del producto o servicio (objeto), se debe resolver una cuestión fundamental: ¿cuál de las tres metodologías básicas del análisis conjunto (tradicional, adaptativa o aditiva) debería utilizarse? La elección de metodologías conjuntas gira en torno a tres características básicas de la investigación propuesta: número de atributos manejados, nivel del análisis y forma del modelo permitida. La Tabla 7.5 compara las tres metodologías a partir de estas consideraciones. El **análisis conjunto tradicional**, representado en el ejemplo anterior, se caracteriza por un modelo aditivo simple que contiene nueve factores estimados para cada individuo. Aunque éste ha sido el pilar de los estudios de análisis conjunto durante muchos años, se han desarrollado dos metodologías adicionales en un intento por tratar con ciertos aspectos del diseño. El método **adaptativo conjunto** se desarrolló para dar cabida a un gran número de factores (muchas veces más de 30) que no sería factible en un análisis conjunto tradicional.

El método **basado en la elección** no sólo emplea una forma única de presentar los estímulos en conjunto en lugar de uno a uno, sino que también difiere en que incluye directamente interacciones y debe ser estimado a nivel agregado. Muchas veces los objetivos del investigador crean situaciones que no se pueden manejar bien con el análisis conjunto tradicional, pero en las que pueden emplearse estas metodologías alternativas. Los problemas de establecer el número de atributos y de seleccionar la forma del modelo se discuten con mayor detalle en la siguiente sección, que se centra en el análisis conjunto tradicional. A continuación, en las secciones ulteriores, se tratan las características específicas de las otras dos metodologías. El investigador debe tener en cuenta que los temas básicos que se discuten en esta sección también se aplican a las otras dos metodologías.

TABLA 7.5. Una comparación de metodologías conjuntas alternativas.

	Metodología conjunta		
Características	*Tradicional*	*Adaptativo*	*Basado en elección*
Número máximo de atributos	9	30	6
Nivel de análisis	Individual	Individual	Agregado
Forma del modelo	Aditivo	Aditivo	Aditivo + efectos interación

Diseño de los estímulos

Los fundamentos experimentales del análisis conjunto dan una gran importancia al diseño de los estímulos que van a ser evaluados por los encuestados. El diseño comprende la especificación del valor teórico conjunto mediante la selección de los factores y niveles que se van a incluir en la construcción de los estímulos. Al definir los factores y los niveles, se deben tratar los supuestos para relacionar el carácter general de cada medida, mientras que el resto de las consideraciones son específicas a los factores y niveles. Estos asuntos del diseño son importantes porque afectan a la efectividad de los estímulos propuestos, a la precisión de los resultados y a la relevancia práctica en última instancia.

Características generales de factores y niveles

Antes de discutir los temas específicos relacionados con los factores o niveles, deben tratarse las características aplicables a la especificación tanto de los factores como de los niveles. Cuando hacemos operativos los factores o niveles, el investigador debe asegurarse que las medidas son comunicables y prácticas.

Medidas comunicables En primer lugar, los factores y niveles deben ser fácilmente comunicables para una evaluación realista. Los métodos tradicionales de administración (papel y lápiz o computador) limitan los tipos de factores que se pueden incluir. Por ejemplo, es difícil describir la fragancia real de un perfume o la sensación de una loción de manos. Las descripciones escritas no recogen bien los efectos sensoriales a menos que los encuestados tengan el producto a mano, huelan la fragancia o utilicen la loción. En un intento por obtener una representación más realista de las características sensoriales que podrían haber sido excluidas en el pasado, se han desarrollado formas específicas de análisis conjunto que emplean la realidad virtual [48] o un abanico completo de efectos sensoriales y multimedia para describir un producto o servicio [37, 54].

Medidas prácticas Los factores o niveles deben ser capaces de ponerse en práctica, lo que significa que los atributos deben ser distintos y representar un concepto que se puede implementar de forma precisa. No deben ser atributos vagos como la calidad o la conveniencia total. Además, los niveles no deberían especificarse en términos imprecisos tales como bajo, alto o moderado. Tales especificaciones son imprecisas por las diferencias entre la percepción de los individuos y lo que realmente quieren decir (por ejemplo, un precio puede ser alto para una persona y bajo para otra). Además, en ocasiones, los conceptos no son fáciles de poner en práctica; por tanto, el investigador no está seguro de si el producto o servicio finalmente desarrollado es en realidad lo que estaba siendo evaluado por el encuestado. Si estos factores poco claros no pueden ser definidos más precisamente, se puede utilizar un proceso de dos etapas. Un estudio preliminar conjunto define qué determina los juicios de esos factores (calidad o conveniencia). A continuación, los factores identificados como importantes en el estudio preliminar se incluyen en un estudio de más envergadura en términos más precisos.

Especificación de supuestos en relación con los factores

Una vez que se han seleccionado los atributos que se van incluir como factores y se ha asegurado que las medidas serán comunicables y prácticas, el investigador debe tratar tres temas específicos de la definición de factores: el número de factores a incluir, la multicolinealidad entre los factores y el papel único del precio como factor.

Número de factores El número de factores incluidos en el análisis afecta directamente a la eficiencia estadística y a la fiabilidad de los resultados. A medida que se añaden más factores y niveles, el creciente número de parámetros a estimar exige o bien un número mayor de estímulos o

bien una reducción de la fiabilidad de los parámetros. El número mínimo de estímulos que deben ser evaluados por un encuestado si el análisis se realiza a nivel individual es:

$$\text{Número mínimo de estímulos} = \text{número total de niveles para todos los factores} - \text{número de factores} + 1$$

Por ejemplo, un análisis conjunto con cinco factores y con tres niveles cada uno (un total de 15 niveles) necesitaría un mínimo de once (15 – 5 + 1) estímulos. Este problema es similar al que nos encontramos en la regresión cuando el número de observaciones era insuficiente para estimar coeficientes válidos. Es especialmente importante en el análisis conjunto porque cada encuestado genera el número exigido de observaciones y por tanto, el problema no puede «resolverse» añadiendo más encuestados.

Multicolinealidad entre factores La multicolinealidad entre los factores es un problema que debe ser solucionado. La correlación entre los factores (conocida como **correlación del entorno** o **inter-atributos**) denota una falta de independencia conceptual entre los factores. En tales casos, los parámetros estimados se ven afectados igual que en la regresión (el Capítulo 4 tiene una discusión de la multicolinealidad y de su impacto). Además, la colinealidad inter-atributos termina generalmente en combinaciones no creíbles de dos o más factores. Por ejemplo, la potencia en caballos y el consumo de gasolina se supone que están negativamente correlacionados. Como resultado, ¿en qué medida es creíble un automóvil con los mayores niveles tanto de potencia como de gasto de gasolina? El problema no reside en los niveles en sí mismos sino en el hecho de que no pueden ser asociados de forma realista en todas las combinaciones, aspecto que se requiere para estimar los parámetros.

Si la multicolinealidad crea estímulos irreales, el investigador tiene dos opciones. La más directa es crear «superatributos» que combinen los aspectos de los atributos correlacionados. En nuestro ejemplo de la potencia y el gasto en gasolina, quizá se puede sustituir por un factor de «rendimiento». Como ejemplo de atributos positivamente correlacionados, factores de diseño de la tienda, luminosidad y decoración se pueden tratar mejor con «atmósfera de la tienda». En todos los casos, cuando se añaden esos superatributos, deben ser tan específicos y prácticos como sea posible. Si no es posible definirlos en términos más amplios, entonces el investigador se verá obligado a eliminar uno de ellos.

La segunda opción implica dos modificaciones posibles a la metodología subyacente del análisis conjunto. La primera modificación comprende diseños experimentales refinados y técnicas de estimación, que crean estímulos cuasi-ortogonales, que pueden ser utilizados para eliminar cualquier estímulo resultante de una correlación inter-atributos [61]. La segunda modificación es restringir la estimación de los componentes parciales de la utilidad total. Estas restricciones pueden ser entre factores así como relacionadas con los niveles dentro de cualquier factor aislado [59, 64]. Cualquiera de estas dos modificaciones a la metodología debería ser considerada sólo después de que se hayan valorado las soluciones más directas, debido a que añaden una considerable complejidad al diseño y estimación del análisis conjunto.

El papel único que representa el precio como factor El precio es un factor que está incluido en muchos estudios conjuntos, puesto que representa un componente distinto de valor para muchos productos o servicios a estudiar. El precio, sin embargo, no es un factor más en su relación con otros factores [33]. En muchos casos, si no la mayoría, el precio tiene un alto grado de correlación inter-atributos con otros factores. Para muchos atributos, un aumento en la cantidad del atributo está asociado con un aumento en el precio, y un nivel de precios descendente puede considerarse irreal. En segundo lugar, la relación calidad-precio puede estar operando entre ciertos factores, de tal forma que ciertas combinaciones pueden ser irreales. En tercer lugar, pueden estar incluidos otros muchos factores «positivos» (calidad, fiabilidad) en la definición de la utilidad del producto o servicio. Sin embargo, cuando se define lo que se «renuncia» por esa utilidad (es decir, el precio), sólo se incluye un factor. Inherentemente esto puede disminuir la importancia del precio. Finalmente, el precio puede entrar en interacción con otros factores, particularmente con

factores más intangibles como el nombre de la marca. El impacto de una interacción en esta situación es que un cierto nivel del precio tiene diferentes significados para diferentes marcas —en un caso puede ser un «premio» de marca y en otro un «descuento» de marca». Discutiremos posteriormente en este capítulo los conceptos de interacción.

Todas estas características singulares del precio como factor no deberían provocar un rechazo del uso de precio por parte del investigador, sino que deberían intentar anticipar los impactos y ajustar el diseño y la interpretación exigida. En primer lugar, se han desarrollado modelizaciones específicas del análisis conjunto, tales como el análisis de valor conjunto (CVA) para ocasiones en las cuales el centro de atención es el precio [53]. Además, si se consideran importantes las interacciones del precio con otros factores, los métodos basados en la elección o los multietapas [47] ofrecen relaciones cuantitativas de estas relaciones. Incluso si no se hace un ajuste específico, el investigador debería considerar estos supuestos en la definición de niveles de precios y en la interpretación de los resultados.

Especificación de supuestos en relación con los niveles

La definición de los niveles es un aspecto crítico del análisis conjunto debido a que los niveles son las medidas reales utilizadas para formar los estímulos. Por tanto, además de ser prácticos y comunicables, la investigación ha demostrado que el número de niveles, el equilibrio entre los niveles y los factores y el rango de los factores tienen distintos efectos sobre las evaluaciones de los encuestados.

Número equilibrado de niveles Los investigadores deberían intentar siempre que sea posible equilibrar o igualar el número de niveles para todos los factores. Se ha encontrado que la importancia relativa estimada de una variable aumenta a medida que el número de niveles lo hace, incluso si los extremos siguen siendo los mismos [68, 69]. Se ha conjeturado que la categorización muy desagregada llama la atención sobre el atributo y provoca que los consumidores se centren en unos factores más que en otros. Si se conoce la importancia relativa a priori de los factores, entonces el investigador puede desear aumentar los niveles de los factores más importantes para evitar una disolución de la importancia y obtener información adicional sobre los factores más importantes [67].

Rango de los niveles de un factor El rango (amplio o reducido) de los niveles debería fijarse de alguna forma fuera de los valores existentes pero no en niveles improbables. Con esto, se reduce la correlación inter-atributos, pero también se puede reducir la credibilidad, por lo que los niveles no deberían ser muy extremos. Unos niveles completamente inaceptables pueden también provocar problemas sustanciales y deberían eliminarse. Antes de excluir un nivel, sin embargo, el investigador debe asegurar que es verdaderamente inaceptable, dado que muchas veces las personas seleccionan productos o servicios que tienen lo que ellos denominan niveles inaceptables. Si se encuentra un nivel inaceptable después de haber realizado el experimento, las soluciones recomendadas son o bien eliminar todos los estímulos que hacen los niveles inaceptables o reducir las estimaciones de los componentes parciales de la utilidad total del nivel excesivo a un nivel tan bajo que cualquier objeto que contenga ese nivel no sea elegido.

El investigador también debe aplicar los criterios de relevancia práctica y factibilidad en la definición de los niveles. Niveles que son impracticables o que nunca se utilizarían en situaciones reales pueden aceptar artificialmente a los resultados. Por ejemplo, supongamos que en el curso normal de la actividad de negocios, el rango de precios varía un 10 por ciento alrededor de un precio medio de mercado. Si se incluyera un nivel de precio por debajo del 20 por ciento, pero que no se fuera a ofrecer en realidad, su inclusión distorsionaría marcadamente los resultados. Los encuestados lógicamente serían más favorables hacia ese nivel de precios. Cuando se realizan estimaciones de los componentes parciales de la utilidad total y se calcula la importancia de los precios, los precios aparecerán artificialmente como más importantes de lo que serían en la realidad de las decisiones del día a día. El investigador debe aplicar los criterios de factibilidad y relevancia práctica para todos los niveles de atributos para asegurar que los estímulos no son creados,

de tal forma que se viesen de forma favorable por el encuestado pero que nunca tendrían la posibilidad real de ocurrir.

Especificación de la forma básica del modelo

Para que el análisis conjunto explique la estructura de preferencias del encuestado sólo a partir de las evaluaciones conjuntas de un conjunto de estímulos, el investigador debe tomar dos decisiones claves en relación con el modelo conjunto subyacente. Estas decisiones afectan tanto al diseño de los estímulos como al análisis de las evaluaciones del encuestado.

La regla de composición: Seleccionar un modelo aditivo frente a uno interactivo

La decisión de más alcance a que tiene que enfrentarse el investigador implica la especificación de la **regla de composición** del encuestado. La regla de composición describe cómo combina el encuestado los componentes parciales de la utilidad total de los factores para obtener el valor conjunto.

El modelo aditivo La regla de composición más común, simple y básica es un **modelo aditivo**, con el cual el encuestado simplemente «suma» los valores de cada atributo (los componentes parciales de la utilidad total) para conseguir el valor total de una combinación de atributos (productos o servicios). Por ejemplo, supongamos que un producto tiene dos factores con componentes parciales de 3 y 4. Por tanto, la utilidad total sería simplemente 7. El modelo aditivo tiene en cuenta la mayoría (entre el 80 o el 90 por ciento) de la variación de la preferencia en casi todos los casos, y es suficiente para la mayoría de las aplicaciones. Es también el modelo básico subyacente tanto en el análisis conjunto tradicional como en el adaptativo (véase Tabla 7.5).

Incorporación de los efectos interacción La regla de composición que utiliza los **efectos interacción** es similar a la forma aditiva, puesto que supone que el consumidor suma los componentes parciales de la utilidad total para todo el conjunto de atributos. Difiere en que permite que ciertas combinaciones de niveles sean superiores o inferiores a la suma. Utilizando nuestro ejemplo previo, un modelo interactivo permitiría que la suma de los dos niveles sea superior o inferior a 7, el resultado del modelo aditivo. En nuestro ejemplo del limpiador industrial, un encuestado puede realmente preferir una cierta marca (genérico), pero sólo con cierto tipo de ingrediente (con fosfatos). En este caso, la marca tiene un componente parcial de utilidad bajo excepto cuando se combina con otro nivel específico (con fosfatos) de los ingredientes de limpieza. Decimos qué marca e ingredientes están interactuando y utilizamos los efectos aditivos para cada factor. La forma interactiva se corresponde con este enunciado, «el conjunto es mayor (o menor) que la suma de sus partes.»

Muchas veces, la incorporación de términos de interacción al modelo disminuye el poder predictivo porque la reducción de la eficacia estadística (más estimaciones de componentes parciales) no se ve compensada por aumentos en el poder predictivo ganados con las interacciones. Las interacciones predicen una varianza sustancialmente menor que los efectos aditivos, y a menudo no exceden de un 5 a un 10 por ciento de aumento de la varianza explicada. Es más probable que los términos de interacción sean relevantes en casos en los que los atributos son menos tangibles, particularmente cuando las reacciones estéticas y emocionales juegan un papel importante. Por ejemplo, hemos discutido antes los efectos interacción que tienen lugar muchas veces entre precio y nombre de marca, que es menos tangible, pero que genera percepciones concretas. La creciente importancia de los términos de interacción proviene de la incapacidad para representar las diferencias efectivas entre ciertos atributos, al estar las partes «inexplicadas» asociadas sólo a ciertos niveles de un atributo.

Un ejemplo de efectos interacción sobre estimaciones de componentes parciales de la utilidad total.

Volviendo a nuestro ejemplo anterior del limpiador industrial, podemos plantear una situación donde las interacciones parecen influir en las elecciones. Supongamos que un tercer encuestado hace el siguiente orden de preferencias:

		Forma	
Marca	*Ingredientes*	*Liquido*	*Polvo*
HATCO	Sin fosfatos	1	2
	Con fosfatos	3	4
Genérico	Sin fosfatos	7	8
	Con fosfatos	5	6

Estos rangos se formaron suponiendo que este encuestado prefiere HATCO y normalmente prefiere limpiadores sin fosfatos a limpiadores con fosfatos. Sin embargo, una mala experiencia con un limpiador genérico hace que el encuestado seleccione con fosfatos en lugar de sin fosfatos si se trata de una marca genérica. Esta elección se denomina efecto interacción entre los factores de marca e ingredientes. Si consideramos sólo un modelo aditivo, obtenemos los siguientes coeficientes:

Forma		*Ingredientes*		*Marca*	
Liquido	*Polvo*	*Sin fosfatos*	*Con fosfatos*	*HATCO*	*Genérico*
0,42	−0,42	0,0	0,0	1,68	1,68

Utilizando estos coeficientes para calcular los órdenes de preferencia de las combinaciones obtenemos lo siguiente:

Calificación real	1	2	3	4	5	6	7	8
Prevista	1,5	3,5	1,5	3,5	5,5	7,5	5,5	7,5

Las predicciones son obviamente menos precisas, dado que sabemos que existe la interacción. También, los coeficientes están equivocados por que los efectos principales de marca e ingredientes están confundidos por las interacciones. Si procediéramos sólo con un modelo aditivo, estaríamos violando uno de los supuestos principales y realizando predicciones potencialmente imprecisas.

Examinar las interacciones de primer orden es una tarea razonablemente simple. Para los datos de preferencia previos, podemos formar tres matrices de orden de preferencia de doble entrada. Por ejemplo, en la primera matriz, se enumeran los dos órdenes de preferencias (uno para cada marca) para cada combinación de forma e ingredientes y después se suman. Para comprobar las interacciones, se suman a continuación los valores de la diagonal y se calcula la diferencia. Si el total es cero, entonces no existe interacción. Como se ve en el ejemplo, la única interacción encontrada es entre marca e ingredientes. A medida que la diferencia se hace mayor, el impacto de la interacción aumenta, y depende del investigador decidir cuándo plantean las interacciones suficientes problemas en la predicción para justificar el aumento de la complejidad de los coeficientes a estimar para los efectos interacción.

Seleccionar el tipo de modelo La elección de una regla de composición determina el tipo y número de tratamientos o estímulos que el encuestado debe evaluar, junto con la forma del método de estimación utilizado. Una forma aditiva exige menor número de evaluaciones por parte del encuestado, y es más fácil obtener estimaciones para los componentes parciales de la utilidad to-

tal. Sin embargo, la forma interactiva puede ser una representación más precisa de cómo los encuestados valoran realmente un producto o servicio. El investigador no conoce con certeza la mejor forma del modelo, pero debe entender las implicaciones de cada elección tanto en el diseño del estudio como de los resultados obtenidos. Si se selecciona una forma de modelo aditiva, no es posible estimar los efectos interacción. Esto no quiere decir que el investigador deba incluir siempre los efectos interacción, en la medida en que añaden una mayor complejidad al proceso de estimación y, en la mayoría de los casos, provocan que el análisis se desarrolle a un nivel agregado en lugar de a un nivel individual. Examinaremos la necesidad de hacer esta elección y de los equilibrios asociados junto con la elección de la forma en varios momentos de nuestra discusión.

Selección de las relaciones de los componentes parciales de la utilidad total: Lineal, cuadrática o componentes parciales separados

La flexibilidad del análisis conjunto en el tratamiento de diferentes tipos de variables viene de los supuestos que hace el investigador en relación con las relaciones de los componentes parciales de la utilidad total con un factor. Al tomar decisiones sobre la regla de composición, el investigador decide qué factores están relacionados con otros en el proceso de decisión del encuestado. Al definir el tipo de relación de los componentes parciales de la utilidad total, el investigador se centra en cómo están relacionados los niveles de un factor.

Tipos de relaciones de los componentes parciales de la utilidad total El análisis conjunto ofrece al investigador tres alternativas, que van desde la más restrictiva (una relación lineal) a la menos restrictiva (componentes parciales de la utilidad total separados), con el punto ideal, o modelo cuadrático, entre ambas alternativas. La Figura 7.2 ilustra las diferencias entre los tres tipos de relaciones. El modelo lineal es el más simple aunque el más restrictivo, debido a que estima un único componente (de la misma forma que un coeficiente de regresión), que se multiplica por el valor del nivel para llevar a valores de los componentes parciales separados para cada nivel. En la forma cuadrática, también conocida como modelo ideal, se relaja el supuesto de estricta linealidad, de tal forma que tenemos una relación curvilínea. La curva puede ser hacia arriba o hacia abajo. Finalmente, la alternativa de los componentes aislados (a menudo denominada simplemente de los componentes parciales) es la más general, al permitir estimaciones aisladas para cada nivel. Cuando se usa este último método, el número de valores estimado es el más alto y se incrementa rápidamente a medida que añadimos más factores y niveles, porque cada nivel tiene una estimación propia.

La forma de relación de los componentes parciales de la utilidad total puede especificarse para cada factor separadamente, e incluso es posible una mezcla de formas si es necesario. Esta elección no afecta a cómo se crean los estímulos o los tratamientos, ya que se siguen calculando los componentes parciales para cada nivel. Afecta, sin embargo, a cómo y a qué tipos de componentes parciales de la utilidad total son estimados por el análisis conjunto. Si se especifican formas li-

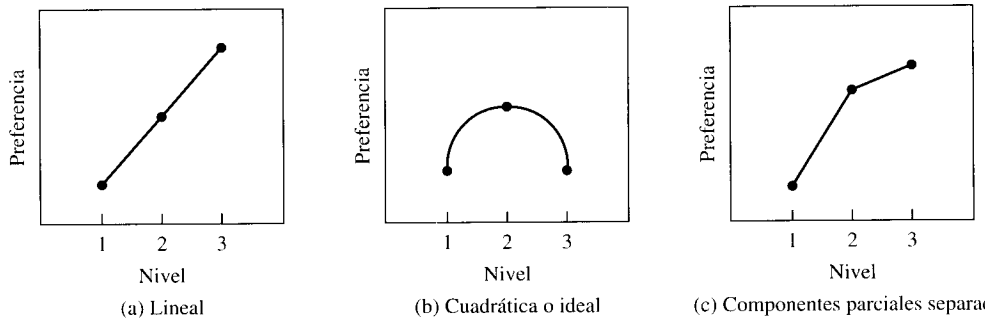


FIGURA 7.2. Los tres tipos de relaciones básicas entre los niveles de los factores en el análisis conjunto.

neales o cuadráticas, entonces los valores de los componentes parciales de la utilidad total para cada nivel se estiman a partir de la relación, con estimaciones de los componentes hechas por separado. Si podemos reducir el número de componentes parciales de utilidad total estimados por cualquier conjunto de estímulos, utilizando un número más restringido de relaciones de los componentes parciales de la utilidad total (es decir, forma lineal o cuadrática), los cálculos serán más efectivos y factibles desde el punto de vista de la estimación estadística. Pero debemos considerar el equilibrio entre estas ganancias y la posibilidad de una representación más precisa de cómo el consumidor forma realmente las preferencias conjuntas si empleamos relaciones de componentes parciales de la utilidad total menos restrictivas.

Selección de una relación de componentes parciales de la utilidad total El investigador tiene varias formas de decidir sobre el tipo de relación para cada factor. En primer lugar, puede basarse en modelos conceptuales o de investigación previos para establecer el tipo de relación. En una aproximación empírica, el modelo conjunto puede ser estimado previamente como un modelo de componentes parciales de la utilidad total, y las diferentes estimaciones de los componentes de la utilidad total pueden examinarse visualmente para detectar si es más apropiada una forma lineal o una forma cuadrática. En muchos casos, la forma general es aparente, y el modelo puede ser reestimado con relaciones especificadas para cada variable de forma justificada. Alternativamente, puede evaluar los cambios en la capacidad predictiva bajo diferentes combinaciones de relaciones para una o más variables, pero esto no se recomienda sin al menos cierta evidencia empírica o teórica de los posibles tipos de relaciones consideradas (es decir, estimaciones previas de los componentes parciales de la utilidad total). Sin ese apoyo, los resultados pueden tener una alta capacidad predictiva pero escaso interés en la toma de decisiones. En cualquier caso, el investigador debe sopesar la capacidad predictiva con el uso que se pretende dar al estudio, el trasfondo conceptual disponible y el grado de relevancia práctica e interpretación necesaria.

Recogida de datos

Una vez que se han especificado los factores y los niveles, además de la forma básica del modelo, el investigador decide a continuación el tipo de representación de los estímulos (*trade-off*, perfil completo o comparación pareada), el tipo de variable respuesta y el método de recogida de datos. El objetivo es transmitir al encuestado las combinaciones de atributos (estímulos) de la forma más realista y eficiente posible. A menudo, los estímulos se presentan en descripciones escritas, aunque los modelos físicos y psíquicos pueden ser muy útiles para atributos estéticos o sensoriales.

Elegir un método de presentación

Los métodos de ***trade-off***, **perfil completo** y **comparación pareada** son las tres técnicas de presentación de los estímulos más frecuentes en el análisis conjunto. Aunque difieren notablemente en la forma y en la cantidad de información presentada al encuestado (véase Figura 7.3), todos ellos son aceptables dentro del análisis conjunto tradicional. La elección entre los métodos de presentación se centra en los supuestos acerca de la cantidad de datos que se están procesando durante el desarrollo del análisis conjunto y el tipo de estimación que se está empleando.

Método de presentación trade-off El método trade-off compara dos atributos al mismo tiempo mediante la clasificación de todas las combinaciones de niveles (Véase Figura 7.3). Tiene la ventaja de ser sencillo y fácil para el encuestado y evita la sobrecarga de información al presentar sólo dos atributos al mismo tiempo. Sin embargo, el uso de este método ha disminuido drásticamente en los últimos años debido a varias limitaciones: (1) un sacrificio en el realismo debido al uso de sólo dos factores al mismo tiempo, (2) el gran número de juicios necesarios incluso para un número reducido de niveles, (3) una tendencia a confundir a los encuestados o seguir un tipo de respuesta rutinaria a causa de la fatiga, (4) la incapacidad de emplear estímulos gráficos o no literarios,

(5) el uso exclusivo de respuestas no métricas y (7) su incapacidad para utilizar los diseños de estímulos de factorial fraccionado para reducir el número de comparaciones realizadas. Recientes estudios han mostrado que la tercera aproximación, las comparaciones pareadas, han desplazado al método de *trade-off* a un segundo lugar en aplicaciones comerciales [69].

Método de presentación de perfil completo El método de presentación más habitual es el de perfil completo, principalmente por su realismo en la percepción y su capacidad para reducir el número de comparaciones a través del uso de diseños factoriales fraccionales. En esta aproximación, cada estímulo se describe por separado, a menudo en una tarjeta de perfiles (véase Figura 7.3 como ejemplo). Esta aproximación obtiene pocos juicios, pero cada uno es más complejo y los juicios pueden ser clasificados o calificados. Entre sus ventajas están (1) una descripción más realista conseguida por la definición de un estímulo en términos de un nivel para cada factor, (2) un retrato más explícito de los *trade-off* entre *todos* los factores y las correlaciones ambientales existentes entre los atributos, y (3) el posible uso de más tipos de juicios de preferencia, tales como las intenciones de compra, probabilidad del juicio y oportunidades de dar marcha atrás –todas difíciles de realizar con el método del *trade-off*.

El método de perfil completo no es perfecto y tiene dos limitaciones principales. En primer lugar, a medida que el número de factores aumenta, también lo hace la posibilidad de sobrecarga de información. El encuestado puede verse tentado de simplificar el proceso centrándose en sólo unos

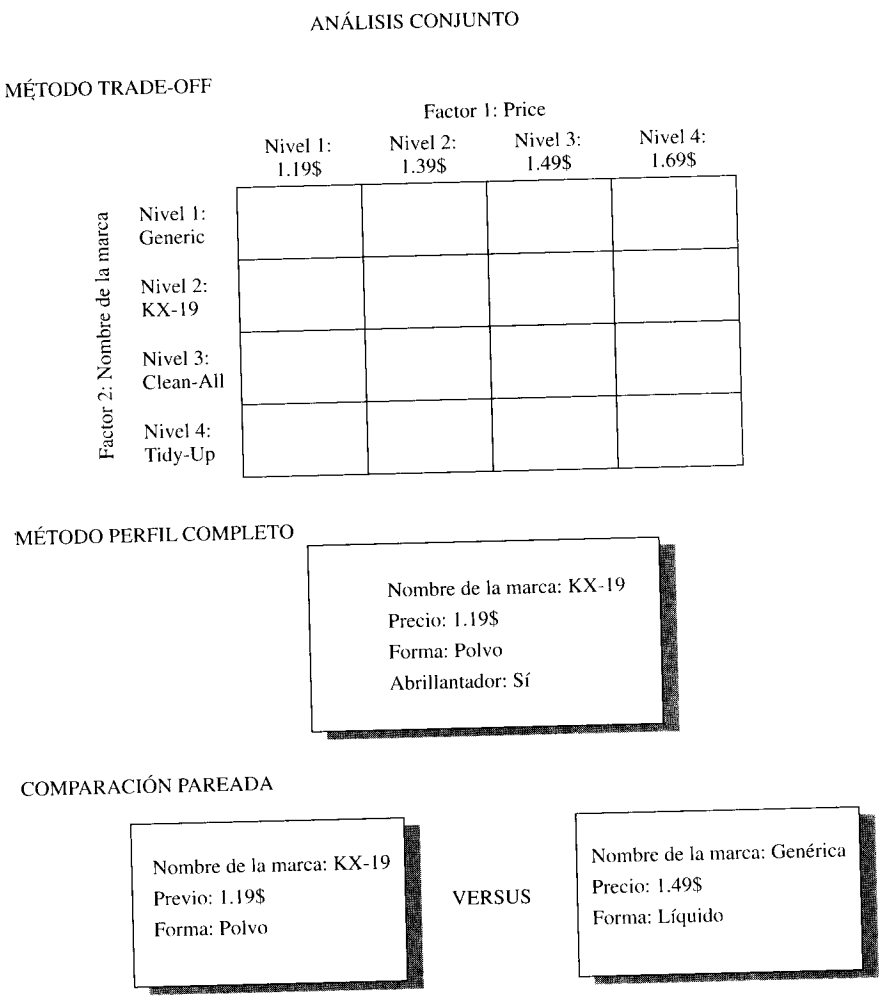


FIGURA 7.3. Ejemplos de los métodos de presentación de los estímulos, según los tres métodos: Trade-off, perfil completo y comparación pareada.

pocos factores, cuando en una situación real se considerarían todos los factores. En segundo lugar, el orden en que se presenta la relación de los factores en la tarjeta de estímulos puede tener su impacto en la evaluación. Por tanto, el investigador necesita alternar los factores entre los encuestados cuando sea posible para minimizar los efectos del orden. Se recomienda el método de perfil completo cuando el número de factores no sea superior a seis. Cuando el número de factores varía de siete a diez, entonces el método *trade-off* se convierte en una posible solución respecto al método de perfil completo. Si el número de factores supera los 10, entonces se sugiere un método alternativo (conjunto adaptativo) [20].

El método de presentación de combinaciones pareadas Este tercer método de presentación combina los dos anteriores. La combinación pareada es una comparación de dos perfiles (véase Figura 7.3), utilizando a menudo el encuestado una escala de calificación para indicar la fuerza de la preferencia por un perfil sobre otro [31]. La característica distintiva de la combinación pareada es que el perfil normalmente no contiene todos los atributos, como ocurre en el método de perfil completo, sino que sólo se seleccionan en un momento unos pocos atributos en la construcción de los perfiles. Es similar al método de *trade-off* en que los pares se evalúan, pero en el caso del método de *trade-off* los pares evaluados son atributos, mientras que en el método de comparación pareada los pares son perfiles con múltiples atributos. El método de comparación pareada es también instrumental en muchos diseños conjuntos especializados, tales como el análisis conjunto adaptativo (ACA) [51], que se utiliza en conjunción con un mayor número de atributos (posteriormente se realizará una discusión más detallada acerca del tratamiento de un mayor número de atributos).

Creación de los estímulos

Una vez que se han seleccionado los factores y los niveles y se ha elegido el método de presentación, el investigador vuelve a la tarea de crear los tratamientos o estímulos a evaluar por los encuestados. Para cualquier método de presentación, el investigador siempre se enfrenta a un aumento de la carga de información al encuestado a medida que el número de factores y niveles aumenta. El investigador debe ponderar los beneficios de aumentar el esfuerzo frente a la información adicional generada. Las siguientes secciones detallan los supuestos implícitos en la creación de estímulos para cada método de presentación.

El método de presentación de trade-off En el caso del método de *trade-off*, se utilizan todas las posibles combinaciones de atributos. El número de matrices de *trade-off* se basa estrictamente en el número de factores y se puede calcular como:

$$\text{Número de matrices de } \textit{trade-off} = \frac{N(N-1)}{2}$$

donde N es el número de factores. Por ejemplo, cinco factores resultarían en 10 matrices de *trade-off* ($5 \times 4/2 = 10$). El investigador debería recordar, sin embargo, que cada matriz de *trade-off* comprende un número de respuestas igual al producto de los niveles de los factores. Por ejemplo, una matriz *trade-off* con factores de 3 niveles cada uno exige nueve evaluaciones (3×3) en cada matriz aislada. Si los cinco factores de nuestro ejemplo tienen cada uno tres niveles, entonces el encuestado evaluaría 10 matrices de *trade-off*, cada una con 9 evaluaciones, para una total de 90 evaluaciones conjuntas. Como podemos ver, este método de presentación puede llevar rápidamente a sobrecargar excesivamente de información al encuestado a medida que el número de atributos o niveles aumenta. Sin embargo, este método sigue haciendo las cosas sencillas al preguntar al encuestado que evalúe sólo dos factores al mismo tiempo, mientras que los otros métodos de presentación pueden llegar a verse bastante comprometidos por la complejidad de los estímulos.

Los métodos de presentación de combinación pareada o perfil completo Los dos métodos restantes —perfil completo y combinación pareada— comprenden la evaluación de un estímulo cada vez (perfil completo) o pares de estímulos (comparación pareada). En un análisis

conjunto simple con un reducido número de factores y niveles (tales como los discutidos previamente para tres factores con dos niveles cada uno que resultaban en ocho combinaciones), el encuestado puede evaluar todos los posibles estímulos. A esto se le conoce como un **diseño factorial** cuando se utilizan todas las combinaciones. Pero a medida que aumenta el número de factores y niveles, el diseño se hace impracticable de forma similar a la mostrada por el método de *trade-off*. Si el investigador está interesado en evaluar el impacto de cuatro variables con cuatro niveles para cada variable, se crearían 256 estímulos (4 niveles × 4niveles × 4 niveles × 4 niveles) en un diseño factorial completo para el método de perfil completo. Evidentemente esto es demasiado para que lo evalúe un solo encuestado y además dar contestaciones consistentes y significativas. Se crearía un número incluso superior de pares de estímulos para las combinaciones pareadas de perfiles con diferentes números de atributos. Lo que se necesita es un método de desarrollo de un conjunto de los estímulos totales que pueda ser evaluado y siga ofreciendo la información necesaria para hacer predicciones precisas y fiables de los componentes parciales de la utilidad total.

Definición de conjuntos de estímulos Un **diseño factorial fraccionado** es el método más común de definición de un conjunto de estímulos a evaluar. El diseño factorial fraccionado selecciona una muestra de posibles estímulos, donde el número de estímulos depende del tipo de regla de composición que se supone que usan los encuestados. Utilizando el modelo aditivo, que supone sólo los efectos principales para cada factor sin interacciones, un estudio que utilice el método de perfil completo con cuatro factores a cuatro niveles requiere sólo de 16 estímulos para estimar los efectos principales. La Tabla 7.6 muestra dos posibles conjuntos de 16 estímulos. Los 16 estímulos pueden ser cuidadosamente construidos para asegurar la correcta estimación de los efectos principales. Los dos diseños de la Tabla 7.6 son **diseños óptimos**, en la medida en que son **ortogonales** (no existe correlación entre los niveles y atributos) y **equilibrados** (cada nivel aparece en el factor el mismo número de veces).

La creación de un diseño óptimo, con ortogonalidad y equilibrado no significa, sin embargo, que todos los estímulos de ese diseño sean aceptables para ser evaluados. Existen varias razones

TABLA 7.6. Diseños factoriales fraccionados alternativos para un modelo aditivo (sólo efectos principales) con cuatro factores con cuatro niveles cada uno

	Diseño 1: niveles para[a]				*Diseño 2: niveles para*[a]			
Estímulo	*Factor* 1	*Factor* 2	*Factor* 3	*Factor* 4	*Factor* 1	*Factor* 2	*Factor* 3	*Factor* 4
1	3	2	3	1	2	3	1	4
2	3	1	2	4	4	1	2	4
3	2	2	1	2	3	3	2	1
4	4	2	2	3	2	2	4	1
5	1	1	1	1	1	1	1	1
6	4	3	4	1	1	4	4	4
7	1	3	2	2	4	2	1	3
8	2	1	4	3	2	4	2	3
9	2	4	2	1	3	2	3	4
10	3	3	1	3	3	4	1	2
11	1	4	3	3	4	3	4	2
12	3	4	4	2	1	3	3	3
13	1	2	4	4	2	1	3	2
14	2	3	3	4	3	1	4	3
15	4	4	1	4	1	2	2	2
16	4	1	3	2	4	4	3	1

[a] Los números que aparecen en las columnas bajo el factor 1 hasta el 4 son los niveles para cada factor. Por ejemplo, el primer estímulo del diseño 1 consiste del nivel 3 para el factor 1, nivel 2 para el factor 2, nivel 3 para el factor 3 y nivel 1 para el factor 4.

para que produzcan estímulos inaceptables. La primera es la creación de estímulos «obvios» —estímulos cuya evaluación es obvia debido a la combinación de niveles. Los ejemplos más comunes son los estímulos que tienen en todos los niveles los valores más altos o más bajos. En estos casos, los estímulos en realidad proporcionan poca información acerca de la elección y pueden crear una percepción de incredulidad por parte del encuestado. La segunda razón es la creación de estímulos increíbles a causa de la correlación inter-atributos, que pueden crear estímulos con combinaciones de niveles de atributos (alto consumo de gasolina, alta aceleración) que no son realistas. Finalmente, se pueden encontrar restricciones en las combinaciones de atributos. En cualquiera de estos casos, los estímulos inaceptables presentan elecciones increíbles para el encuestado y deberían ser eliminadas para asegurar una estimación válida del proceso así como una percepción de credibilidad de la elección entre los encuestados.

Existen varias alternativas para eliminar los estímulos inaceptables. En primer lugar, el investigador puede generar otro diseño factorial fraccional y evaluar la aceptabilidad de sus estímulos. Si todos los diseños contienen estímulos inaceptables y no se puede encontrar una mejor alternativa de diseño, entonces deberían destruirse los estímulos inaceptables. Aunque el diseño no sea totalmente ortogonal (es decir, que esté de alguna forma correlacionado y se denomina **cuasi-ortogonal**), no se viola ningún supuesto del análisis conjunto. Esto creará problemas similares a los de la multicolinealidad en la regresión (es decir, inestabilidad de las estimaciones cuando se cambian ligeramente los niveles y una menor capacidad para evaluar el impacto aislado de cada atributo). Todo los diseños cuasi ortogonales deberían ser evaluados por su **eficiencia del diseño**, que es una medida de la correspondencia del diseño en términos de ortogonalidad y equilibrio para un diseño óptimo [36]. Utilizando escalas de 100 puntos (un diseño óptimo = 100), se pueden evaluar los diseños no ortogonales alternativos y seleccionar el diseño más eficiente con todos los estímulos aceptables. La mayoría de los programas de análisis conjunto evalúan la eficacia de los diseños para desarrollar diseños cuasi ortogonales.

Las correlaciones inter-atributos se deben tratar en términos conceptuales. Los estímulos inaceptables debidos a las correlaciones inter-atributos pueden darse en diseños óptimos y ortogonales, y el investigador debe acomodarlos dentro del diseño. En términos prácticos, las correlaciones inter-atributos deberían minimizarse pero no hacerse necesariamente cero si correlaciones reducidas (0,20 o menos) añaden realismo. La mayoría de los problemas se encuentran en el caso de correlaciones negativas, como entre el consumo de gasolina y caballos de potencia. La adición de factores no correlacionados puede reducir la correlación media inter-atributos, de tal forma que con un número realista de factores (es decir, 6 factores), la intercorrelación media estaría cercana a 0,20, lo que tendría efectos sin consecuencias. Pero el investigador debería evaluar siempre la credibilidad de un estímulo como medida de relevancia práctica.

Se necesitan los 240 posibles estímulos de nuestro ejemplo que no están incluidos en el diseño factorial fraccional seleccionado si se van a estimar los 11 términos de interacción. Como ya se ha discutido, el investigador puede decidir que las interacciones seleccionadas son importantes y deberían incluirse en la estimación del modelo. En este caso, el diseño factorial fraccionado debería incluir estímulos adicionales para acomodar las interacciones. Las guías publicadas para los diseños factoriales fraccionados o componentes de los programas de análisis conjunto diseñan los subconjuntos de estímulos para mantener la ortogonalidad o eficiencia máxima en diseños «cuasi» ortogonales, haciendo la generación de estímulos de perfil completo bastante fácil [1, 12, 22, 43].

Si el número de factores se hace muy grande y la metodología adaptativa conjunta no es aceptable, se puede emplear un **diseño de puente** [5]. En este diseño, los factores se dividen en subconjuntos de tamaño apropiado, donde algunos atributos se solapan entre los conjuntos de tal forma que cada conjunto tiene un factor(es) en común con otros conjuntos de factores. Los estímulos se construyen para cada subconjunto de tal forma que el encuestado nunca ve el número original de factores en un perfil único. Cuando se estiman los componentes parciales de la utilidad total, se combinan los conjuntos separados de perfiles y se proporciona un único conjunto de estimaciones. Los programas informáticos manejan la división de los atributos, creación de estímulos y su recombinación para la estimación [9]. Cuando se utilizan comparaciones pareadas, el número puede ser muy grande y complejo, de modo que muy a menudo se utilizan los programas informáti-

cos interactivos que seleccionan los conjuntos óptimos de pares a medida que se realiza el cuestionario.

Seleccionar una medida de la preferencia del consumidor

El investigador debe también seleccionar la medida de preferencia: clasificación de orden frente a calificación (es decir, una escala de 1 a 10). Aunque el método *trade-off* emplea sólo datos de clasificación, el método de comparación pareada puede evaluar la preferencia, bien mediante la obtención de una clasificación de preferencia de un estímulo frente a otro o bien como una medida binaria, lo que es preferible. El método de perfil completo también se acomoda tanto al método de clasificación como al método de calificación.

Cada medida de preferencia tiene ciertas ventajas y limitaciones. Obtener una medida de preferencia de clasificación de orden (es decir, ordenación de los estímulos desde el más preferido al menos preferido) tiene dos ventajas principales: (1) es probable que sea más fiable porque la ordenación es más fácil que la calificación para un número razonablemente reducido (menos de 20) de estímulos, y (2) proporciona más flexibilidad en la estimación de los diferentes tipos de reglas de composición. Tiene, sin embargo, una desventaja principal: es difícil de administrar, dado que el proceso de ordenación se realiza normalmente clasificando tarjetas que recogen los estímulos y este procedimiento sólo puede realizarse mediante una entrevista personal.

La alternativa es obtener una calificación de preferencia en una escala métrica. Las medidas métricas se analizan y se administran más fácilmente, incluso por correo, y permiten realizar estimaciones conjuntas mediante regresiones multivariantes. Sin embargo, los encuestados pueden ser menos discriminantes en sus juicios de lo que lo serían con una ordenación. También, dado el elevado número de estímulos evaluados, es útil aumentar el número de categorías de respuesta por encima del considerado en la mayoría de los estudios del consumidor. Como norma general se toman 11 categorías (es decir, calificaciones de 0 a 10 o 0 a 100 en incrementos de 10) para 16 o menos estímulos y se expande a 21 categorías para más de 16 estímulos [38].

Realización del estudio

En el pasado, la complejidad del análisis conjunto ha llevado muy a menudo al uso de entrevistas personales para obtener las respuestas conjuntas. Las entrevistas personales permiten al entrevistador explicar los (a veces) difíciles métodos asociados con el análisis conjunto. Desarrollos recientes en los métodos de entrevista, sin embargo, han hecho factible realizar análisis conjuntos tanto por correo (con cuestionarios de lápiz y papel o por computador) como por teléfono. Si el estudio está diseñado para asegurar que el encuestado puede asimilar y procesar apropiadamente el estímulo, entonces todos los métodos de encuesta arrojan relativamente la misma precisión predictiva [2]. El uso de entrevistas por computador que envían el disquete por correo ha simplificado enormemente las demandas que el análisis conjunto exigía al encuestado y ha hecho muy asequible desarrollos tales como los análisis conjuntos adaptativos [51].

Una de las preocupaciones de cualquier estudio conjunto es la carga que se pone en el encuestado debido al número de estímulos conjuntos evaluados. Obviamente, el encuestado no podría evaluar los 256 estímulos de nuestro anterior diseño factorial, pero ¿cuál es el número apropiado de tareas en un análisis conjunto? Una revisión reciente de los estudios de marketing con análisis conjunto encontró que los encuestados podían completar fácilmente hasta 20 evaluaciones conjuntas [34]. Después de muchas evaluaciones, las respuestas empiezan a ser menos creíbles y menos representativas de la estructura de referencia subyacente. Aunque no existe un número mínimo o máximo absoluto de evaluaciones de estímulos, el investigador siempre debería esforzarse en utilizar los menos posibles a la vez que se mantiene la eficiencia en el proceso de estimación. Tampoco existe sustituto para contrastar previamente un estudio conjunto para evaluar la carga que soporta el encuestado, el método de realización y la aceptabilidad de los estímulos.

Tercer paso: Supuestos básicos del análisis conjunto

El análisis conjunto tiene el grupo de supuestos menos restrictivo en la estimación del modelo conjunto. El diseño experimental estructurado y la naturaleza generalizada del modelo hace innecesarios la mayoría de los tests realizados en otros métodos de dependencia. Por tanto, los tests estadísticos de normalidad, homocedasticidad e independencia que se desarrollaron en otras técnicas de dependencia no son necesarios. El uso de diseños de estímulos con base estadística asegura también que la estimación no está confundida y que los resultados son interpretables bajo la regla de composición asumida.

Así, incluso aunque hay pocos supuestos estadísticos, los supuestos conceptuales son quizá mayores que con cualquier otra técnica multivariante. Como ya se ha mencionado, el investigador debe especificar la forma general del modelo (efectos principales frente a modelo con interacciones) antes de diseñar la investigación. Con ello, se «construye» esta decisión y se hace imposible contrastar modelos alternativos una vez que la investigación está diseñada y los datos se han recogido. El análisis conjunto no es como la regresión, por ejemplo, donde los efectos adicionales (términos de interacción o no lineales) pueden ser evaluados fácilmente. El investigador debe tomar esta decisión teniendo en cuenta la forma del modelo y debe diseñar la investigación consecuentemente. Por tanto, el análisis conjunto, aunque tiene pocos supuestos estadísticos, está muy determinado por la teoría en su diseño, estimación e interpretación.

Cuarto paso: Estimación del modelo conjunto y valoración del ajuste global

Las opciones disponibles para el investigador en términos de técnicas de estimación han aumentado notablemente en los últimos años. El desarrollo de técnicas en conjunción con métodos especializados de presentación de estímulos (es decir, el análisis conjunto adaptativo o basado en elección) es sólo un tipo de mejora de esta clase. El investigador, al obtener los resultados de un estudio de análisis conjunto, sin embargo, debe tratar los supuestos de selección del método de estimación y evaluación de los resultados (véase Figura 7.4).

Selección de una técnica de estimación

Las evaluaciones de ordenación exigen una forma modificada de análisis de varianza específicamente diseñado para datos ordinales. Entre los programas informáticos más populares y más conocidos están MONANOVA (análisis monotómico de la varianza) [31, 35] y LINMAP [55]. Estos programas dan estimaciones de los atributos de los componentes parciales de la utilidad total, de tal forma que la ordenación de su suma (valor total) para cada tratamiento esté correlacionada lo más cercanamente posible a la clasificación observada. Si se obtiene una medida métrica de preferencia (es decir, calificaciones en lugar de clasificaciones), entonces muchos métodos, incluido el análisis de regresión, pueden servir para estimar los componentes parciales de la utilidad total para cada nivel. La mayoría de los programas informáticos existentes hoy en día pueden utilizar tanto el tipo de evaluación (calificaciones o clasificaciones), así como estimar cualquiera de los tres tipos de relaciones (lineal, punto ideal y componentes parciales de la utilidad total).

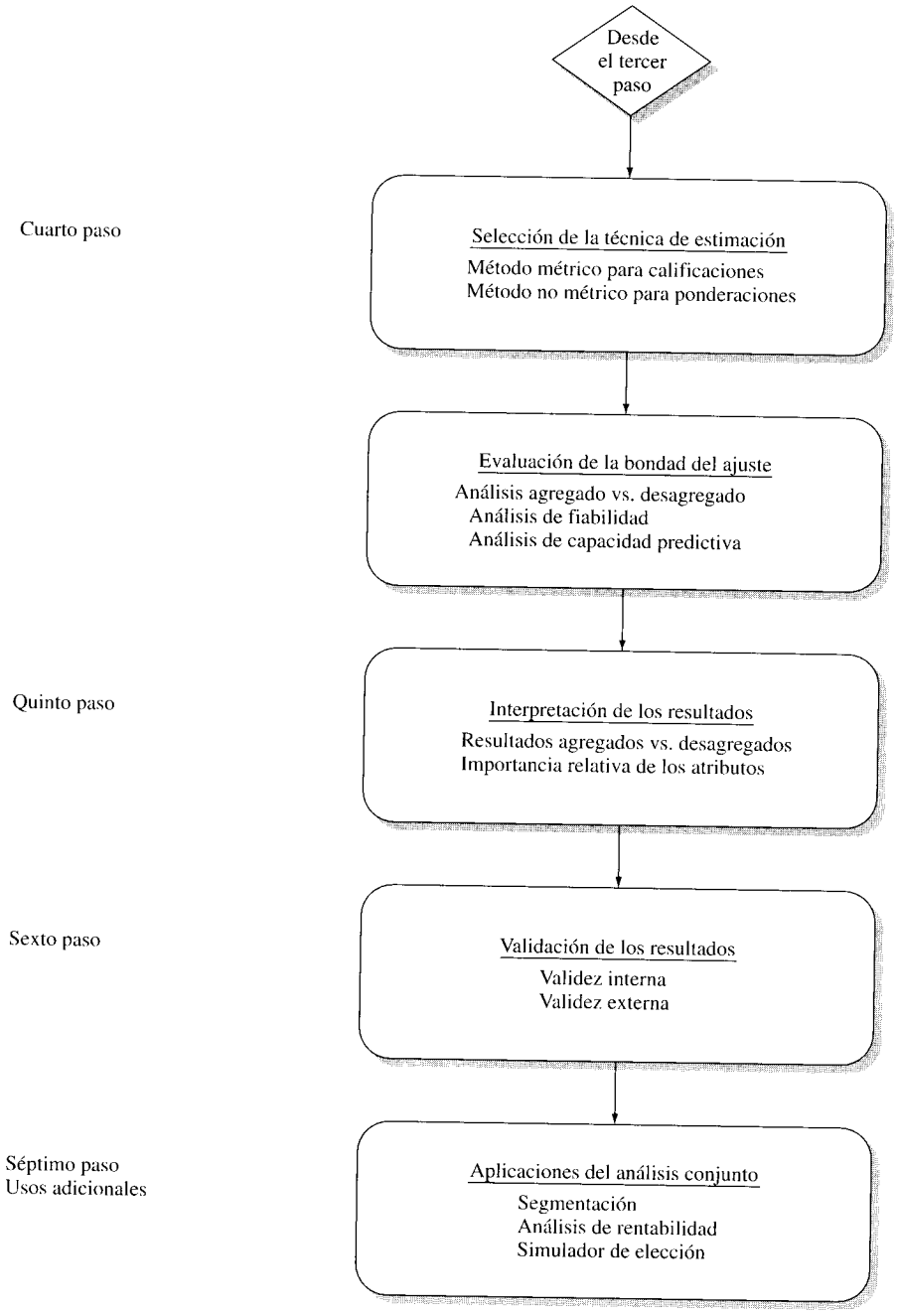


FIGURA 7.4. Pasos 4-7 del diagrama de decisión del análisis conjunto.

Evaluación de la bondad del ajuste del modelo

La precisión de los resultados del análisis conjunto se evalúa tanto a nivel individual como agregado. El objetivo es averiguar la consistencia con que predice el modelo el conjunto de evaluaciones de preferencias dadas por cada persona. Para los datos de clasificación, se utilizan las correlaciones basadas en las clasificaciones previstas y efectivas (es decir, la *ro* de Spearman o la *tau* de Kendall). Si se obtiene una calificación métrica, entonces es apropiada una simple corre-

lación de Pearson, la misma que se utiliza en la regresión, junto con una comparación de las calificaciones previstas y actuales. En casos de predicción a nivel individual, las preferencias previstas y efectivas están correlacionadas para cada persona y contrastadas para la significación estadística.

En la mayoría de los experimentos conjuntos, sin embargo, el número de estímulos no excede sustancialmente del número de parámetros, y siempre existe la posibilidad de «sobreajustar» los datos. Se recomienda encarecidamente a los investigadores que midan la precisión del modelo no sólo sobre los estímulos originales sino también con un conjunto de **estímulos *holdout*** o **de validación**. En un procedimiento similar a la muestra holdout en el análisis discriminante, el investigador prepara más tarjetas de estímulo que las necesarias para la estimación de los componentes parciales de la utilidad total, y los encuestados las califican todas al mismo tiempo. Se utilizan entonces parámetros del modelo conjunto estimado para predecir la preferencia sobre el nuevo conjunto de estímulos, que se compara con las respuestas efectivas para evaluar la fiabilidad del modelo. Los individuos que tienen un ajuste predictivo muy bajo para la muestra holdout se pueden eliminar del análisis. La muestra holdout también ofrece al investigador una oportunidad de evaluación directa de estímulos de interés para el estudio.

Si se utiliza una técnica de estimación agregada, el investigador puede usar una muestra holdout de encuestados en cada grupo para evaluar la precisión predictiva. Este método no es factible para resultados desagregados porque no hay un modelo «generalizado» que se aplique a la muestra holdout, en la medida en que cada encuestado de la muestra de estimación tiene estimaciones de los componentes parciales de la utilidad total individualizadas.

Quinto paso: Interpretación de los resultados

Análisis agregado frente a análisis desagregado

La aproximación normal para interpretar el análisis conjunto es la desagregada; esto es, cada encuestado se modeliza separadamente y los resultados del modelo se examinan para cada encuestado. El método más común de interpretación es un examen de las estimaciones de los componentes parciales para cada factor, evaluando su magnitud y su pauta tanto a efectos de relevancia práctica como a efectos de correspondencia con relaciones teóricas entre niveles. Cuanto mayor sea el componente parcial (positiva o negativa), mayor será el impacto que tenga sobre la utilidad total. Los valores de los componentes parciales de la utilidad total pueden ser representados gráficamente para identificar las pautas. Muchos programas convierten las estimaciones de los componentes parciales de la utilidad total a una escala común (por ejemplo, de un máximo de 100 puntos) para que se puedan realizar comparaciones entre los factores individuales, así como entre individuos. Esta conversión ofrece un medio de utilizar los componentes parciales de la utilidad total en otras técnicas multivariantes tales como el análisis *cluster*.

La interpretación también se puede realizar con resultados agregados. Si la estimación del modelo se hace a nivel individual y después agregado, o se hacen estimaciones agregadas para un conjunto de encuestados, el análisis ajusta un modelo al agregado de respuestas. Como se podría esperar, este proceso generalmente ofrece unos resultados pobres cuando se intenta predecir lo que haría cualquier encuestado aislado o cuando se quieren interpretar los componentes parciales de la utilidad total para cualquier encuestado aislado. A menos que el investigador esté tratando con una población que muestre concluyentemente un comportamiento homogéneo con referencia a los factores, no debería utilizarse el análisis agregado como método de análisis. Sin embargo, muchas veces el análisis agregado predice mejor un comportamiento agregado como la cuota de mercado. Por tanto, el investigador debe identificar el objetivo primordial del estudio y emplear el nivel apropiado de análisis o una combinación de los niveles de análisis.

Evaluación de la importancia relativa de los atributos

Además de representar el impacto de cada nivel con las estimaciones de los componentes parciales de la utilidad total, el análisis conjunto puede evaluar la importancia relativa de cada factor. Dado que las estimaciones de los componentes parciales de la utilidad total se convierten normalmente a una escala común, la mayor contribución a la utilidad conjunta —y por tanto el factor más importante— es el factor con el mayor rango (alto o bajo) de los componentes parciales de la utilidad total. Esto permite la comparación entre los encuestados en una escala común así como dar significado a la magnitud de la puntuación de importancia. El investigador debe considerar siempre el impacto sobre los valores de importancia de un nivel extremo o prácticamente improbable. Si se encuentra un nivel así, debería ser eliminado del análisis o los valores de importancia se verían reducidos a reflejar sólo el rango de los niveles factibles.

Sexto paso: Validación de los resultados del análisis conjunto

Los resultados del análisis conjunto pueden validarse tanto internamente como externamente. La validación interna implica la confirmación de que la regla de composición seleccionada (es decir, aditiva frente a interactiva) es la apropiada. El investigador está normalmente limitado nada más que a la evaluación empírica de la validez de la forma del modelo seleccionado en un estudio completo, debido a las necesidades de datos para contrastar ambos modelos. Esta evaluación se realiza más efectivamente mediante la comparación de modelos alternativos (aditivos frente a interactivos) en un examen previo para confirmar qué modelo es el apropiado. Ya hemos discutido el uso de estímulos holdout para evaluar la precisión predictiva para cada individuo o una muestra holdout de encuestados si el análisis se realiza a nivel agregado.

La validación externa implica en general la capacidad del análisis conjunto de predecir elecciones efectivas, y de forma más específica la representatividad de la muestra. Aunque el análisis conjunto se ha empleado en numerosos estudios en los últimos 20 años, muy poca investigación científica se ha centrado en su validez externa. Un estudio confirmó que el análisis conjunto se corresponde estrechamente con los resultados de la tradicional contrastación de concepto, una metodología aceptada para la predicción de las preferencias del cliente [63]. Aunque no hay evaluación del error de muestreo en los modelos de nivel individual, el investigador debe asegurarse siempre de que la muestra es representativa de la población a estudiar. Esto es especialmente importante cuando los resultados conjuntos se utilizan para la segmentación o simulación de elecciones (véase la siguiente sección para una discusión más detallada de estos usos de los resultados conjuntos).

Aplicación de los resultados del análisis conjunto

Normalmente, los modelos conjuntos estimados a nivel individual (un modelo por individuo) se utilizan en una o más de las siguientes áreas de apoyo a la toma de decisiones mediante la representación de los procesos de decisión de los individuos. Con resultados a nivel individual, se pueden representar empíricamente las preferencias de cada individuo. Los resultados agregados conjuntos pueden representar grupos de individuos y ofrecer también un medio de predecir sus decisiones en cualquier tipo de situación. Las aplicaciones más comunes del análisis conjunto con su representación de la estructura de preferencias del consumidor incluyen la segmentación, análisis de rentabilidad y simuladores conjuntos.

Segmentación

Uno de los usos más comunes de los resultados del análisis conjunto a nivel individual es agrupar a los encuestados con componentes parciales de la utilidad total o valores de importancia similares para identificar los segmentos. Los componentes parciales estimados de la utilidad total pueden utilizarse por separado o en combinación con otras variables (es decir, demográficas) para obtener agrupaciones de encuestados que son más parecidos en sus preferencias [17]. En el ejemplo del limpiador industrial, podemos encontrar un grupo para el que la marca es lo más importante, mientras que otro grupo puede valorar más otros ingredientes. Para el investigador interesado en conocer la presencia de tales grupos y de sus magnitudes relativas, existen varias aproximaciones a la segmentación, todas con diferentes fortalezas y debilidades.

Análisis de rentabilidad

Un complemento a la toma de decisiones del diseño del producto es un análisis de rentabilidad marginal del diseño del producto propuesto. Si se conoce el coste de cada característica, el coste de cada «producto» puede combinarse con la cuota de mercado esperada y el volumen de ventas para predecir su viabilidad. Este proceso puede señalar a una combinación de atributos con una reducida cuota de mercado como la más rentable debido a un aumento del margen de beneficio resultante del bajo coste de determinados componentes. Adjunta al análisis de rentabilidad está la evaluación de la sensibilidad del precio, que puede ser tratada a través de diseños de investigación específicos [47] o programas especializados [53]. Pueden utilizarse en este análisis tanto los resultados individuales como los agregados.

Simuladores conjuntos

Hasta ahora, el investigador todavía entiende sólo la importancia de los atributos y el impacto sobre niveles específicos. Pero, ¿cómo consigue el análisis conjunto sus otros objetivos primarios —utilizando el análisis «qué-pasa-si» para predecir la cuota de preferencias que es probable que capture un estímulo (real o hipotético) en varios escenarios competitivos de interés para la gestión? Éste es el papel que juegan los **simuladores de elección**, que siguen un proceso de tres pasos:

1. Estimar y validar los modelos conjuntos para cada encuestado (o grupo).
2. Seleccionar los conjuntos de estímulos a contrastar de acuerdo a los posibles escenarios competitivos.
3. Simular las elecciones de todos los encuestados (o grupos) para los conjuntos específicos de estímulos y predecir las participaciones en la preferencia de cada estímulo agregando sus elecciones.

Una vez que se ha estimado el modelo conjunto, el investigador puede especificar cualquier número de conjuntos de estímulos para la simulación de las elecciones del consumidor. Entre los usos posibles están: (1) la evaluación del impacto de añadir un producto a un mercado existente; (2) el aumento potencial de una estrategia multimarca o multiproducto, incluyendo estimaciones de canibalismo comercial: o (3) el impacto de eliminar una marca o un producto del mercado. En cada caso, el investigador proporciona el conjunto de estímulos que representan al mercado y se simulan las elecciones de cada encuestado.

Los simuladores de elección utilizan normalmente dos tipos de reglas en la predicción de la elección de un estímulo [16]. La primera es el modelo de máxima utilidad, que asume que el encuestado elige el estímulo con la mayor utilidad prevista. Este modelo es el más apropiado en casos de mercados con individuos de preferencias muy diferentes y en situaciones que comprendan compras esporádicas y no rutinarias. La regla de elección alternativa es una medida de probabilidad de compra, en la que la predicción de la suma de las probabilidades de elección suma el 100 por cien so-

bre el conjunto de estímulos contrastado. Esta aproximación se ajusta a situaciones de compra repetitivas, para las que las compras pueden estar más atadas a situaciones habituales en el tiempo. Los dos métodos más comunes de hacer estas predicciones son los modelos logit y los BTL (Bradford-Terry-Luce), que realizan predicciones bastante parecidas en casi todas las situaciones [24].

Se advierte al investigador, sin embargo, sobre la suposición de que la participación de la preferencia en una simulación conjunta se traslade directamente a la participación en el mercado. La simulación conjunta representa sólo aspectos del producto y quizá del precio interesantes para la gestión de marketing, omitiendo todos aquellos factores del marketing (por ejemplo, publicidad y promoción, distribución y respuestas competitivas) que afectan en última instancia a la cuota de mercado. La simulación conjunta presenta una visión del mercado de producto y la dinámica de preferencias que puede verse en la muestra bajo estudio.

Metodologías alternativas en el análisis conjunto

Hasta este punto hemos tratado con las aplicaciones del análisis conjunto incluidas en la metodología tradicional del análisis conjunto. Pero las aplicaciones del mundo real, en muchas ocasiones implican 20 o 30 atributos o exigen un marco de decisión más realista de lo que hemos utilizado en discusiones anteriores. La investigación reciente se ha dirigido a solucionar estos problemas encontrados en muchos estudios conjuntos, con dos nuevas metodologías que se están desarrollando: (a) un modelo adaptativo conjunto para tratar con un gran número de atributos, y (b) un análisis conjunto basado en la elección que proporciona elecciones más realistas. Estas áreas representan el eje fundamental de la investigación actual del análisis conjunto [11, 20].

Análisis conjunto adaptativo: Análisis conjunto con un gran número de factores

Los métodos de perfil completo y *trade-off* empiezan a ser inmanejables cuando consideran entre 6 y 9 atributos, aunque muchos estudios conjuntos necesitan incorporar entre 20 y 30 atributos. En estos casos, se utiliza una forma adaptada o reducida del análisis conjunto para simplificar el esfuerzo de la recogida de datos y representar una decisión realista. Las dos opciones son los modelos auto-explicados y los modelos adaptativos e híbridos.

Modelos conjuntos auto-explicados

En el **modelo auto-explicado**, el encuestado ofrece una calificación de la atracción de cada nivel de un atributo y a continuación califica la importancia relativa del atributo conjunto. Los componentes parciales de la utilidad total se calculan mediante una combinación de los dos valores [58]. Se trata de un método composicional donde las calificaciones se realizan sobre los componentes de la utilidad en lugar de sobre una preferencia global. Como variante importante del análisis conjunto y más cercana a los modelos multiatributos tradicionales, esta forma de modelización presenta varios temas de interés. En primer lugar, ¿pueden los encuestados evaluar la importancia relativa de los atributos con precisión cuando la investigación muestra que éstos pueden subestimarse en modelos multiatributos dado que los encuestados quieren dar respuestas socialmente deseables? En segundo lugar, las correlaciones inter-atributos juegan un papel importante y provocan sesgos sustanciales en los resultados debido a la «doble contabilidad» de los factores correlacionados. Finalmente, los encuestados nunca realizan una elección (califican el conjunto de combinaciones hipotéticas de atributos), y esta falta de realismo es una limitación crítica de las aplicaciones sobre nuevos productos. La investigación reciente, sin embargo, ha demostrado que este método puede tener una capacidad predictiva idónea cuando se compara con los métodos de aná-

lisis conjunto tradicionales [18]. Por tanto, si el número de factores no puede reducirse a un nivel práctico aceptable en los métodos de análisis conjunto tradicional, entonces un modelo auto-explicado puede ser un método viable alternativo al análisis conjunto tradicional.

Modelos de análisis conjunto híbridos o adaptativos

Una segunda aproximación es el **modelo híbrido** o **adaptativo**, denominado así porque combina los modelos conjuntos de componentes parciales de la utilidad total y los auto-explicados [14-15]. Utiliza valores auto-explicados al crear un conjunto reducido de estímulos (de tres a nueve) seleccionados a partir de un diseño factorial fraccional. Los estímulos se evalúan a continuación de forma similar al análisis conjunto tradicional. Los conjuntos de estímulos difieren entre los encuestados, y aunque cada encuestado evalúa sólo un número reducido, se evalúan colectivamente todos los estímulos por una parte de los encuestados. La aproximación de integrar información del encuestado para simplificar o aumentar los trabajos de elección ha llevado a que recientemente se hayan realizado diversas investigaciones sobre diferentes aspectos del diseño de investigación [3, 30, 60, 64]. Una de las variantes más comunes de esta aproximación es ACA, un programa informático de análisis conjunto desarrollado por Sawtooth Software [51]. ACA emplea calificaciones auto-explicadas para reducir el tamaño del diseño factorial y hacer el proceso más manejable. Su relativa capacidad predictiva se ha mostrado comparable a la del análisis conjunto tradicional, y es una alternativa adecuada cuando el número de atributos es elevado [18, 32, 63, 70].

Cuando nos enfrentamos con un número de factores que no pueden ser admitidos en los métodos conjuntos discutidos hasta este momento, los modelos híbridos o adaptativos y auto-explicados preservan al menos una parte de los principios que subyacen al análisis conjunto. Al comparar estas dos extensiones, los métodos auto-explicados tienen relativamente una baja fiabilidad, aunque existen investigaciones recientes encaminadas a mejorarla. Cuando los modelos híbridos y los auto-explicados se comparan con los métodos de perfil completo, los resultados son mixtos, aunque funciona ligeramente mejor el método híbrido o adaptativo, particularmente el ACA [25]. Aunque se necesita más investigación para confirmar las comparaciones entre los métodos, los estudios empíricos indican que tanto los métodos híbridos o adaptativos como las nuevas formas de modelos auto-explicados ofrecen alternativas viables al análisis tradicional conjunto cuando se trata con un número elevado de factores.

Análisis conjunto basado en elección: Introducir otro toque de realismo

En los últimos años, muchos investigadores del área del análisis conjunto han dirigido sus esfuerzos hacia una nueva metodología conjunta que proporcione una mayor dosis de realismo en la tarea de elegir. Con el imperioso objetivo de entender el proceso de toma de decisiones del encuestado y predecir el comportamiento en el mercado, el análisis conjunto tradicional asume que el juicio, basado en clasificaciones o calificaciones, recoge la elección del encuestado. Algunos investigadores han argumentado que no es el modo más realista de representar el proceso de decisión efectivo, y otros investigadores han señalado la falta de una teoría formal que vincule estos juicios de medida de la elección [39]. Lo que ha surgido es una metodología conjunta alternativa, conocida como análisis conjunto basado en elecciones, con la validez en principio inherente de preguntar al encuestado que elija un estímulo de perfil completo de un conjunto de estímulos conocidos como un **conjunto de elección**. Esta situación es mucho más representativa del proceso real de selección de un producto de entre un conjunto de productos competidores. Además, el análisis conjunto basado en la elección ofrece una opción de *no* elección de ninguno de los estímulos presentados al incluir una opción de no elección en el conjunto de elección. Mientras que el análisis conjunto tradicional supone que las preferencias de los encuestados se asignan siempre entre el conjunto de estímulos, la aproximación basada en la elección permite una contracción del mercado si todas las alternativas del conjunto de elección no son atractivas.

Una ilustración simple de análisis conjunto de perfil completo frente al análisis basado en la elección

Antes de discutir algunos de los detalles más técnicos del análisis conjunto basado en la elección y en qué medida difiere de otras metodologías conjuntas, examinemos un simple ejemplo a efectos ilustrativos. Una compañía de teléfonos celulares desea estimar el mercado potencial de tres opciones de servicio que pueden añadirse a la tarifa del servicio básico de 14,95$ al mes y 0,50$ al minuto por tiempo de llamada:

ICA – factura detallada con un cargo de 2,75$ al mes.
CW – llamada en espera con cargo de 3,50$ al mes.
TWC – llamada a tres con un cargo mensual de 3,50$.

El análisis conjunto tradicional se realiza con estímulos de perfil completo que representan varias combinaciones de servicio, que van desde el servicio base al servicio base con las tres opciones. En la Tabla 7.7 se muestra el conjunto completo de perfiles (diseño factorial). El estímulo 1 representa el servicio base sin opciones, el estímulo 2 es el servicio base más la factura detallada, y así sucesivamente hasta el estímulo 8, que es el servicio base con las tres opciones (factura detallada, llamada en espera y llamada a tres). Se pide al encuestado que clasifique o califique estos ocho perfiles.

En una aproximación basada en la elección, se muestra al encuestado una serie de conjuntos de elección, cada uno teniendo varios estímulos de perfil completo. Un diseño basado en la elección se muestra también en la Tabla 7.7. El primer conjunto de elección consiste en cinco de los estímulos de perfil completo (estímulos 1, 2, 4, 5 y 6) y una opción de «ninguno de éstos». A continuación el encuestado escoge sólo uno de los perfiles del conjunto de elección («más preferido» o «más deseado») o la opción «ninguno de éstos». En la Tabla 7.8 se muestra un ejemplo de la preparación de un conjunto de elección para la elección por parte del encuestado, en particular para el conjunto 6. La preparación de los estímulos y los conjuntos de elección se basa en los principios de diseño experimental [30 y 39] y es el objeto de un considerable esfuerzo investigador para refinar y mejorar la tarea de elegir [3, 11, 27, 30, 47].

TABLA 7.7. Comparación de diseño de estímulos utilizando el análisis conjunto tradicional y el basado en elección

Análisis conjunto tradicional					
	Niveles de los factores[a]			*Análisis basado en elección*	
Estímulos	*ICA*	*CW*	*TWC*	*Conjunto de elección*	*Estímulos del conjunto de elección*[b]
1	0	0	0	1	1, 2, 4, 5, 6 y no elección
2	1	0	0	2	2, 3, 5, 6, 7 y no elección
3	0	1	0	3	1, 3, 4, 6, 7, 8 y no elección
4	0	0	1	4	2, 4, 5, 7, 8 y no elección
5	1	1	0	5	3, 5, 6, 8 y no elección
6	1	0	0	6	4, 6, 7 y no elección
7	0	1	1	7	1, 5, 7, 8 y no elección
8	1	1	1	8	1, 2, 6, 8 y no elección
				9	1, 2, 3, 7 y no elección
				10	2, 3, 4, 8 y no elección
				11	1, 3, 4, 5 y no elección

[a] Niveles: 1 = opción de servicio incluida; 0 = opción de servicio no incluida.
[b] Los estímulos utilizados en los conjuntos de elección son aquellos definidos por el análisis conjunto tradicional.

TABLA 7.8. Ejemplo de un conjunto de elección en un análisis conjunto basado en la elección

¿Qué sistema de teléfono elegiría?			
1	2	3	4
Sistema base a 14,95$/mes y 0,50/$minuto más: ◆TWC—llamada a tres por sólo 3,50$/mes	Sistema base a 14,95$/mes y 0,50$/minuto más: ◆ICA—facturación detallada por sólo 2,75$/mes	Sistema base a 14,95$/mes y 0,50$/minuto más: ◆CW—llamada en espera por sólo 3,50$/mes y ◆TWC—llamada a tres por sólo 3,50$/mes	Ninguno de éstos

El número de perfiles varía a lo largo de los conjuntos de elección. También, el número de elecciones hechas (una elección para cada uno de los 11 conjuntos de elección) es en realidad superior en este caso a lo que se requeriría en el diseño factorial. Pero a medida que el número de factores y niveles aumenta, el diseño basado en la elección requiere considerablemente menos evaluaciones (recordemos que nuestro ejemplo anterior de cuatro factores con cuatro niveles generaba 256 estímulos). Las ventajas de la aproximación basada en la elección consisten en el realismo adicional que introduce y la capacidad para estimar los términos de interacción, lo que no es posible con el análisis conjunto tradicional. Después de que cada encuestado ha escogido un estímulo para cada conjunto de elección, se agregan los datos para todos los encuestados (segmentos u otra agrupación homogénea de encuestados) para estimar los componentes parciales de la utilidad total para cada nivel y los términos de interacción. Desde estos resultados, podemos evaluar las contribuciones de cada factor y la interacción nivel-factor y estimar las cuotas de mercado probables de mercado de perfiles competidores.

Características únicas del análisis conjunto basado en la elección

La naturaleza básica de los modelos conjuntos basados en la elección y sus precedentes en el campo teórico de la integración de la información [38] ha llevado a una perspectiva en cierta forma más técnica de lo que se puede encontrar en otras metodologías conjuntas. Mientras que las otras metodologías están basadas en sólidos principios experimentales y estadísticos, la complejidad adicional tanto en el diseño de estímulos como en la estimación ha sido objeto de una gran cantidad de investigaciones con el fin de desarrollar estas áreas. Gracias a estos esfuerzos, los investigadores tienen ahora una comprensión más clara de los supuestos implícitos en cada etapa. Las siguientes secciones detallan algunas de las áreas y temas en los cuales el análisis conjunto basado en la elección es distinto del resto de las metodologías conjuntas.

Tipo de proceso de toma de decisiones representada El análisis tradicional conjunto ha estado asociado siempre con una aproximación de información intensiva al proceso de toma de decisiones, caracterizado por el escrutinio de los estímulos de perfil completo compuesto de los niveles de cada atributo. Cada atributo está representado por igual y considerado en un perfil único. Pero en el análisis conjunto basado en la elección, los investigadores llegan a la conclusión de que la tarea de elegir puede implicar un tipo diferente de proceso de toma de decisiones. Al hacer elecciones entre estímulos, los consumidores parecen elegir entre un conjunto reducido de factores sobre los cuales se hacen las comparaciones y, en última instancia, eligen [26]. Esto es similar a los tipos de decisiones asociadas con la restricción del tiempo o estrategias simplificadoras, caracterizadas por un bajo nivel de reflexión. Por tanto, cada metodología conjunta ofrece diferentes visiones en los procesos de toma de decisiones. Dado que los investigadores pueden no estar dispuestos a elegir una sola metodología en este momento, una nueva estrategia es emplear ambas metodologías y deducir perspectivas únicas de cada una de ellas [26].

Diseño de estímulos Quizá la mayor ventaja del modelo conjunto basado en la elección está en el realismo del proceso de elección representado por el conjunto de elección. Dos desarrollos han mejorado aún más la tarea de elegir. En primer lugar, existe una elección más realista e informativa entre alternativas parecidas y comparables, en vez de la situación en la que uno de los estímulos es claramente inferior o superior. Sin embargo, el proceso de diseño de estímulos se centra en conseguir la ortogonalidad y el equilibrio entre los atributos. Un reciente trabajo ha mostrado cómo puede crearse el conjunto de elección para asegurar el equilibrio no entre los niveles de los factores sino entre las utilidades de los estímulos [27]. Esto proporciona un esquema más realista, que puede aumentar la implicación del consumidor y ofrecer mejores resultados. En segundo lugar, en un método que implica información adicional acerca de los encuestados, se crea un conjunto de elección que ajuste las preferencias específicas de cada individuo y conseguir una mejor precisión predictiva en las situaciones basadas en el mercado [9]. Estos dos desarrollos son característicos de los esfuerzos que intentan mejorar la tarea de elegir haciendo de ella un método de evaluación incluso más realista y eficiente de las preferencias del consumidor.

Técnicas de estimación El fundamento conceptual del análisis conjunto basado en la elección es psicológico [40, 62], pero fue el desarrollo de una técnica de estimación logit multinomial [42], que ofrece un método operativo de estimación de estos tipos de modelos de elección. Aunque se han hecho esfuerzos considerables para refinar y hacer más asequible la técnica, todavía representa una metodología más compleja que aquellas asociadas con las otras metodologías conjuntas. Un aspecto particular que todavía queda sin resolver es la propiedad de IIA (independencia de alternativas irrelevantes), un supuesto que hace problemática la predicción de alternativas muy similares. Aunque explorar todos los supuestos que implica IIA está más allá del objeto de esta discusión, se aconseja al investigador que cuando utilice el análisis conjunto basado en la elección entienda las ramificaciones de este supuesto.

Ventajas y limitaciones del análisis conjunto basado en la elección

La creciente aceptación del análisis conjunto entre los practicantes de la investigación de mercados se debe fundamentalmente a la creencia de que la obtención de preferencias, teniendo los encuestados que elegir un único estímulo preferido entre un conjunto de estímulos, es más realista —y por tanto un método mejor— para aproximarse al proceso de decisión efectivo. Además de añadir realismo a la tarea de elección el investigador se encuentra con un número de *«trade-offs»* (equilibrios) que el investigador debe considerar antes de elegir el modelo conjunto basado en la elección.

La tarea de elegir Cada conjunto de elección contiene varios estímulos y cada estímulo contiene varios factores a diferentes niveles, igual que los estímulos de perfil completo. Por tanto, el encuestado debe procesar una cantidad considerablemente mayor de información que otras metodologías del análisis conjunto a la hora de tomar una decisión en cada conjunto de elección. Sawtooth Software, promotores del sistema basado en la elección (CBC), creen que aquellas elecciones que implican más de seis atributos es probable que confundan y abrumen al encuestado [52]. Aunque el método basado en la elección reproduce las decisiones reales, la inclusión de muchas alternativas implica una tarea formidable que normalmente termina con menos información de la que se habría ganado a través de la calificación de los estímulos individualmente.

Precisión predictiva

En la práctica, las tres metodologías conjuntas permiten tipos parecidos de análisis, simulaciones e información, incluso aunque los procesos de estimación sean diferentes. Aunque los modelos basados en la elección todavía tienen que someterse a más contrastaciones empíricas, algunos in-

vestigadores creen que tienen ventaja en la predicción de comportamiento decisor. Sin embargo, las contrastaciones empíricas indican una escasa diferencia entre los modelos basados en la calificación de niveles ajustados con la posibilidad de no elección y los modelos basados en la elección de tipo logit multinominal generalizado [44]. La conclusión fue que tanto el modelo basado en calificaciones como el modelo basado en la elección predicen igualmente bien. Otra investigación comparó las dos aproximaciones al análisis conjunto (basado en las calificaciones o basado en la elección) en términos de la capacidad de predecir participaciones en una muestra holdout [13]. Ambas aproximaciones predicen las elecciones de la muestra holdout bien, sin que ninguna aproximación sea dominante y los resultados se mezclen en diferentes situaciones. En última instancia, la decisión de utilizar un método sobre el otro está dictada por los objetivos y el alcance del estudio, la familiaridad del investigador con cada método y el software disponible para analizar apropiadamente los datos.

Aplicaciones prácticas Los modelos basados en la elección estimados a nivel agregado ofrecen los valores y significación estadística de todas las estimaciones, producen fácilmente predicciones de la cuota de mercado realistas para estímulos nuevos [30], y ofrecen seguros adicionales de que se utilizan «elecciones» entre estímulos para calibrar el modelo. Sin embargo, los modelos de elección agregados dificultan la segmentación del mercado. El análisis conjunto basado en la elección no es capaz de estimar un modelo conjunto diferenciado para cada encuestado, haciendo imposible agrupar a los encuestados de acuerdo a los resultados de sus modelos conjuntos tal y como se ha discutido previamente. Como contraste, los modelos basados en calificación descritos previamente se ajustan bien a los estudios de segmentación pero se enfrentan al problema de tests incómodos de significación estadística de las estimaciones de los componentes parciales de la utilidad total. Por tanto, los resultados pueden ser difíciles de resumir y la simulación de las participaciones de la elección puede ser problemática.

Disponibilidad de los programas informáticos Existen varios programas informáticos basados en la elección a disposición de los investigadores que asisten en todas las fases del diseño de investigación, estimación del modelo e interpretación [29, 52]. Sin embargo, la investigación académica y aplicada se integra lentamente en esos programas disponibles comercialmente. La mayoría de los avances todavía se encuentran limitados a un reducido dominio y no están disponibles para un uso más amplio. Estas mejoras y nuevas capacidades, después de una rigurosa validación por la comunidad académica, deberían convertirse en parte habitual de todos los programas basados en la elección.

Resumen

El análisis conjunto basado en la elección es una metodología de análisis conjunto emergente que sigue siendo prometedora en el crecimiento de la capacidad interpretativa y predictiva del análisis conjunto. El amplio interés y la investigación en la mejora de casi todas las áreas de la metodología proporcionarán los fundamentos necesarios para su crecimiento continuo, disponibilidad y aceptación de este método. Añadirá un componente diferenciado a la «caja de herramientas» del investigador a la hora de entender las preferencias del consumidor.

Resumen de las tres metodologías conjuntas

El análisis conjunto ha evolucionado desde los orígenes que ahora conocemos como análisis conjunto tradicional para desarrollar dos metodologías adicionales, que se centran en dos temas esenciales —tratar con un gran número de atributos y hacer la tarea de elegir más realista. Cada metodología tiene características distintas que ayudan a definir aquellas situaciones en las cuales es más aplicable (véase nuestra anterior discusión del paso 2). Además, en muchas situaciones son

factibles dos o más metodologías y el investigador tiene la opción de elegir una o, cada vez más, combinar las metodologías. Sólo siendo conscientes acerca de los puntos fuertes y de las debilidades de cada metodología los investigadores pueden realizar la elección más apropiada. Se anima a los investigadores interesados en el análisis conjunto a continuar observando los desarrollos de esta técnica multivariante tan ampliamente utilizada.

Una ilustración del análisis conjunto

En las siguientes secciones examinamos los pasos de una aplicación del análisis conjunto al ejemplo del limpiador industrial. La discusión sigue el proceso de construcción de modelos introducido en el Capítulo 1, centrándose en (1) diseño de los estímulos, (2) estimación e interpretación de los componentes parciales de la utilidad total y (3) aplicación de un simulador conjunto para predecir participaciones de mercado para un producto de nueva formulación. Se utiliza la opción *CATEGORIES* del *SPSS*, en las fases de diseño, análisis y simulador de elecciones de este ejemplo [57]. Se obtienen resultados comparables con los otros programas de análisis conjunto para el uso comercial y académico.

Primer paso: Objetivos del análisis conjunto

Al desarrollar un nuevo limpiador industrial, HATCO necesita entender más profundamente las necesidades y preferencias de sus clientes industriales. En un estudio adjunto al descrito en el Capítulo 1, HATCO encarga un experimento de análisis conjunto entre 100 clientes industriales. La investigación de mercados y la consulta con el grupo de desarrollo del producto identificaron cinco factores como atributos determinantes en el segmento objetivo del mercado de limpieza industrial. Los cinco atributos se muestran en la Tabla 7.9.

Segundo paso: Diseño del análisis conjunto

Las decisiones en esta fase son: (1) seleccionar la metodología conjunta a utilizar, (2) diseñar los estímulos a evaluar, (3) especificación de la forma del modelo básico y (4) seleccionar el método de recogida de datos.

TABLA 7.9. Atributos y niveles para el experimento de análisis conjunto de HATCO

Descripción	*Nivel*
Forma del producto	1. Líquido premezclado 2. Líquido concentrado 3. Polvo
Número de aplicaciones por envase	1. 50 2. 100 3. 200
Adición de desinfectante al limpiador	1. Sí 2. No
Limpiador en formulación biodegradable	1. No 2. Sí
Precio normal por pedido	1. 35 centavos 2. 49 centavos 3. 79 centavos

Selección de una metodología conjunta

El primer asunto a resolver es la selección de la metodología conjunta de entre tres opciones —conjunto tradicional, conjunto adaptativo o conjunto basado en elección. Dado el reducido número de factores (cinco), las tres metodologías serían apropiadas. Dado que el interés se centraba en un profundo entendimiento de la estructura de preferencias, fue elegida la metodología tradicional conjunta como más adecuada en términos de carga de respuestas sobre el encuestado y la profundidad de la información representada. Se consideró también el análisis conjunto basado en la elección, pero la ausencia de interacciones propuestas y el deseo de reducir la complejidad de la tarea llevó a la selección de los métodos conjuntos tradicionales.

Diseño de estímulos

La investigación del grupo de seguimiento estableció niveles específicos para cada atributo (véase Tabla 7.2) que se consideraron prácticos y comunicables a través de un estudio previo de evaluación y comprobación a pequeña escala. Los niveles de precios se tomaron del rango encontrado en los productos existentes para asegurar que no fueran irreales. El tipo de producto no sugiere factores intangibles que contribuirían a la correlación inter-atributos y los atributos se definieron específicamente para minimizar las correlaciones inter-atributos.

Especificación de la forma del modelo básico

Después de una consideración cuidadosa, los investigadores de HATCO se sintieron confiados como para asumir que una regla de composición aditiva era la apropiada. Aunque la investigación ha mostrado que a menudo el precio tiene interacciones con otros factores, se supuso que el resto de los factores eran razonablemente ortogonales y que no se necesitaban los términos de interacción. Para permitir que los resultados más generales fueran aplicables a un amplio abanico de posibilidades, se estimaron todos los factores con estimaciones particulares de los componentes parciales de la utilidad total para cada nivel.

Selección del método de recogida de datos

Para asegurar el realismo y permitir el uso de calificaciones en lugar de clasificaciones, HATCO decidió utilizar el método de perfil completo para obtener las evaluaciones de los encuestados. Al elegir una regla aditiva, también podían utilizar un diseño factorial fraccionado para evitar la evaluación de las 108 combinaciones posibles ($3 \times 3 \times 2 \times 2 \times 3$). El componente del diseño del estímulo de los programas informáticos generó un conjunto de 18 descripciones de perfil completo (véase Tabla 7.10), que permitieron la estimación de los principales efectos ortogonales para cada factor. Se generaron cuatro estímulos adicionales para que sirvieran como estímulos de validación.

El experimento de análisis conjunto fue realizado durante una entrevista personal. Después de recoger algunos datos preliminares, se entregó a los encuestados un conjunto de 22 tarjetas, cada una conteniendo descripciones del estímulo de perfil completo. Además fueron presentadas plegadas, con siete categorías de respuestas, que van de «absolutamente improbable que lo compre» a «seguro que lo compro». Se dieron instrucciones a los encuestados para que situasen cada tarjeta en la categoría de respuesta que mejor describiese sus intenciones de compra. Después de colocar las tarjetas, se les pidió que revisaran las colocaciones y reorganizasen las tarjetas si fuera necesario. Los estímulos de validación fueron calificados al mismo tiempo que otros estímulos pero se ocultaron en el análisis en la fase de estimación. Una vez terminado, el entrevistador registró la categoría para cada tarjeta y continuó con la entrevista.

TABLA 7.10. Conjunto de 18 estímulos de perfil completo utilizadas en el experimento de análisis conjunto

Número tarjeta	*Forma del producto*	*Numero de aplicaciones*	*Calidad de desinfectante*	*Biodegradabilidad*	*Precio por pedido*
Estímulos utilizados en la estimación de los componentes parciales de la utilidad total					
1	Concentrado	200	Sí	No	35 centavos
2	Polvo	200	Sí	No	35 centavos
3	Premezclado	100	Sí	Sí	49 centavos
4	Polvo	200	Sí	Sí	49 centavos
5	Polvo	50	Sí	No	79 centavos
6	Concentrado	200	No	Sí	79 centavos
7	Premezclado	100	Sí	No	79 centavos
8	Premezclado	200	Sí	No	49 centavos
9	Polvo	100	No	No	49 centavos
10	Concentrado	50	Sí	No	49 centavos
11	Polvo	100	No	No	35 centavos
12	Concentrado	100	Sí	No	79 centavos
13	Premezclado	200	No	No	79 centavos
14	Premezclado	50	Sí	No	35 centavos
15	Concentrado	100	Sí	Sí	35 centavos
16	Premezclado	50	No	Sí	35 centavos
17	Concentrado	50	No	No	49 centavos
18	Polvo	50	Sí	Sí	79 centavos
Estímulos de validación holdout					
19	Concentrado	100	Sí	No	49 centavos
20	Polvo	100	No	Sí	35 centavos
21	Polvo	200	Sí	Sí	79 centavos
22	Concentrado	50	No	Sí	35 centavos

Tercer paso: Supuestos básicos del análisis conjunto

El supuesto relevante del análisis conjunto es la especificación de la regla de composición y, por tanto, de la forma del modelo utilizada para estimar los resultados conjuntos. En esta situación, la naturaleza del producto, la tangibilidad de los atributos y la falta de intangibilidad o apelaciones emocionales justifica el uso de un modelo aditivo. HATCO confió en la utilización de un modelo aditivo para esta situación de toma de decisiones industrial. Además, simplifica el diseño de los estímulos y facilita los esfuerzos de recogida de datos.

Cuarto paso: Estimación del modelo conjunto y valoración del ajuste del modelo

La estimación de los componentes parciales de la utilidad total y la importancia relativa de cada atributo se hicieron en primer lugar para cada encuestado por separado, y los resultados se agregaron después para obtener un resultado conjunto. Inicialmente, se hicieron estimaciones de los componentes parciales aislados para todos los niveles, examinando las estimaciones individuales emprendidas para examinar la posibilidad de introducir restricciones respecto a la forma de relación del factor (es decir, emplear una forma de relación lineal o cuadrática). La Tabla 7.11 muestra los resultados para la muestra conjunta, así como para cinco encuestados. El examen de los resultados conjuntos sugiere que quizá se podría estimar una relación lineal para la variable precio. Sin embargo, la revisión de los resultados individuales muestra que sólo dos de los cinco encuestados tienen estimaciones de los componentes parciales de la utilidad total para los factores

TABLE 7.11. Resultados del experimento de análisisconjunto para la muestra conjunta y una selección de encuestados

Estimaciones de los componentes parciales de la utilidad total																		
Forma de producto			*Número de unidades*			*Desinfectante*		*Biodegradable*		*Precio por unidad*			*Factores de importancia relativa*					*Precisión predictiva*[a]	
Premezclado	*Concentrado*	*Polvo*	50	100	200	*Sí*	*No*	*No*	*Sí*	*0,35$*	*0,49$*	*0,79$*	1	2	3	4	5	*Estimación*	*Holdout*
Muestra total																			
−0,311	0,192	0,139	−0,416	0,021	0,396	0,486	−0,486	−0,077	0,077	0,962	0,069	−1,031	11,75	18,23	21,82	3,44	44,76	0,988	0,667
Encuestados seleccionados																			
−0,056	0,611	−0,556	0,444	0,611	−1,056	−0,208	0,208	0,542	−0,542	1,444	0,944	−2,389	14,29	20,41	5,10	13,27	46,94	0,929	0,707
−0,444	0,389	0,056	−0,944	−0,278	1,222	1,083	−1,083	−0,417	0,417	0,389	0,889	−1,278	10,20	26,53	26,53	10,20	26,53	0,949	0,707
−1,111	1,222	−0,111	−0,611	0,556	0,056	0,083	−0,083	−0,417	0,417	0,556	1,056	−1,611	32,56	16,28	2,33	11,63	37,21	0,776	0,183
−1,722	0,944	0,778	−1,056	−0,056	1,111	0,792	−0,792	−0,083	0,083	1,944	−0,222	−1,722	26,02	21,14	15,45	1,63	35,77	0,907	0,913
−0,611	0,389	0,222	−0,444	0,222	0,222	−0,417	0,417	−0,542	0,542	2,556	0,056	−2,611	11,43	7,62	9,52	12,38	59,05	0,851	0,707

[a] Se presenta la *ro* de Spearman para la estimación de la precisión predictiva, y la *tau* de Kendall para la precisión predictiva de la muestra holdout.

de precio que sigan en términos generales una pauta lineal. Por tanto, la aplicación de una forma lineal para el factor precio distorsionaría gravemente la relación entre los niveles, por lo que la estimación de componentes parciales de la utilidad total estaba plenamente justificada.

Las medidas de la precisión predictiva para la estimación de los componentes parciales de la utilidad total estaban todas en un rango aceptable tanto para los resultados agregados como para los cinco individuos (véase Tabla 7.11). Aunque los estímulos holdout tenían generalmente un nivel más bajo de precisión predictiva, también estos valores eran aceptables. Por tanto, el investigador retendría a todos estos encuestados por ser adecuados para su representación mediante los resultados del análisis conjunto.

Quinto paso: Interpretación de los resultados

La Figura 7.5 muestra una diversidad de estimaciones de los componentes parciales de la utilidad total para cada factor. Por ejemplo, los resultados conjuntos muestran un impacto escaso de la característica biodegradable. Sin embargo, cuando se muestran los resultados individuales, vemos que el encuestado 2 valora la característica biodegradable, mientras que el encuestado 1 no. También se observan resultados dispares para la característica desinfectante y el número de dosis por envase. Estas diferencias ilustran el valor del análisis conjunto desarrollado por cada encuestado en lugar de basarse exclusivamente en resultados agregados.

La Figura 7.6 compara los valores de importancia obtenidos para cada factor tanto para los resultados agregados como para los resultados desagregados de dos de los encuestados. Aunque observamos una consistencia general en los resultados, cada encuestado tiene aspectos únicos que los diferencian del resto y de los resultados agregados. Las mayores diferencias se observan en las características de biodegradabilidad y calidad desinfectante. Ambos encuestados dan más importancia a la biodegradabilidad de lo que se refleja en los resultados agregados. El encuestado 1 tiene menos interés en la característica desinfectante, mientras que el encuestado 2 está ligeramente por encima de la media en este aspecto. Finalmente, el encuestado 2 es mucho menos sensible al precio

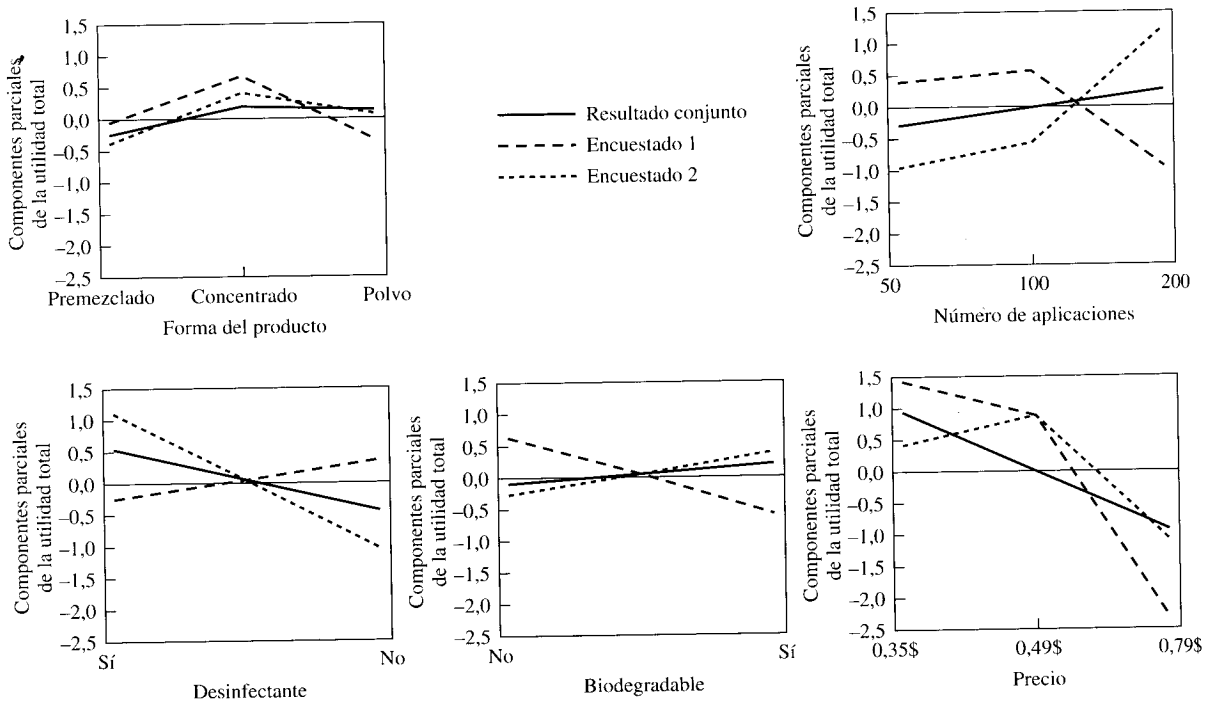


FIGURA 7.5. Estimaciones de los componentes parciales de la utilidad total para los resultados agregados y los encuestados elegidos.

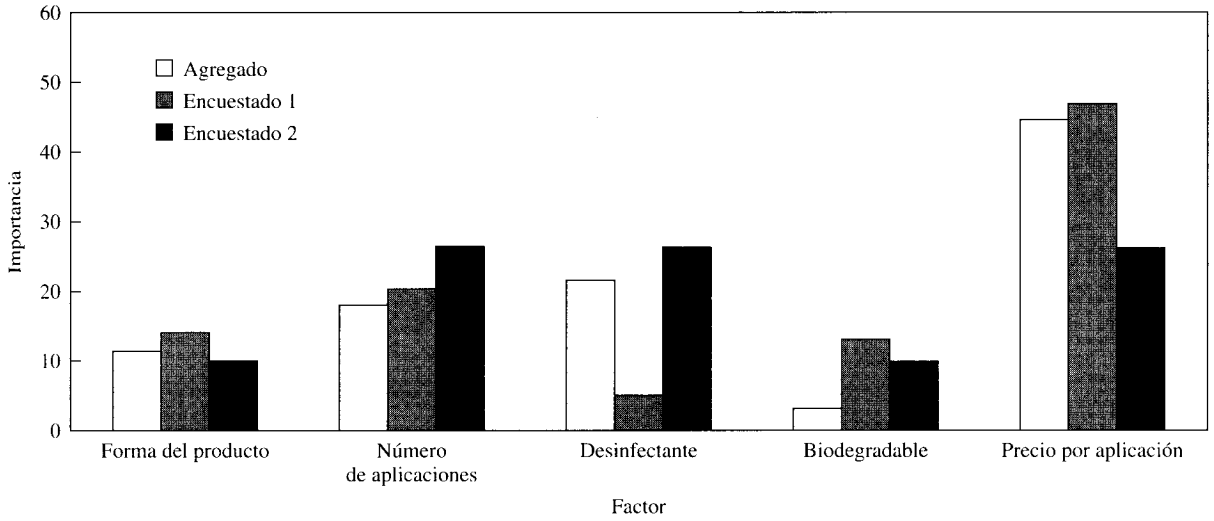


FIGURA 7.6. Importancia del factor para los resultados agregados y los encuestados seleccionados.

de lo que reflejan los resultados agregados o el encuestado 1. En resumen, los resultados muestran dos individuos distintos, lo que ilustra los beneficios de los resultados del análisis conjunto estimados para cada individuo en lugar de para el agregado. Una extensión del análisis conjunto es definir los grupos de encuestados con estimaciones de componentes parciales de utilidad total similares o valores de importancia de los factores utilizando el análisis *cluster*. Estos segmentos pueden ser evaluados y perfilados en sus estructuras específicas de preferencias y potencial de mercado.

Sexto paso: Validación de los resultados

Los altos niveles de precisión predictiva tanto de la estimación como de los estímulos *holdout* entre los encuestados confirman la regla de composición aditiva para este conjunto de encuestados. El punto más importante de la validación es la representatividad de la muestra. En esta situación, lo más probable es que HATCO procediera con un proyecto a gran escala con una mayor cobertura de sus bases de clientes para asegurar la representatividad. Otra consideración es la inclusión de no clientes, especialmente si el objetivo es entender el conjunto de mercado, no sólo a los clientes de HATCO.

Una aplicación práctica: Utilización de un simulador de elección

Además de entender la estructura de preferencias individual y agregada de los encuestados, HATCO utiliza también los resultados del análisis conjunto para simular las elecciones entre tres posibles productos. Los tres productos contrastados fueron los siguientes:

Producto 1. Un limpiador premezclado en un tamaño listo para usar (50 aplicaciones por envase) que no dañaba el medio ambiente (biodegradable) y que cumplía todos los requisitos sanitarios (desinfectante) por sólo 79 centavos la aplicación.

Producto 2. Una versión industrial del producto 1 con las características ambientales y sanitarias señaladas, pero en formato concentrado en grandes envases (200 aplicaciones) a un precio menor por aplicación, 49 centavos.

Producto 3. Un auténtico limpiador en polvo y en tamaño económico (200 aplicaciones por envase) al precio más bajo posible, 35 centavos por aplicación.

El simulador de elección calculó a continuación las estimaciones de las preferencias de los productos para cada encuestado. Las predicciones de las participaciones esperadas de mercado se hicieron con dos modelos de elección: el modelo de máxima utilidad y un modelo probabilístico. El modelo de máxima utilidad cuenta el número de veces que cada uno de los tres productos tiene la utilidad más alta para todos los encuestados. Como se ve en la Tabla 7.12, el producto 1 era preferido (tenía el mayor valor de preferencia previsto) por sólo el 5 por ciento de los encuestados. El producto 2 estaba, a continuación, preferido por el 25,5 por ciento y el más preferido era el producto 3, con 69,5 por ciento. Los porcentajes decimales se deben al ajuste de predicciones entre los productos 2 y 3.

TABLA 7.12. Resultados del simulador de elecciones para las formulaciones de tres productos

		Predicciones de cuota de mercado		
			Modelos probabilísticos	
Formulación del producto	*Puntuaciones de preferencia agregada previstas*	*Modelo de máxima utilidad (%)*	*BTL (%)*	*Logit (%)*
1	2,4	5,0	18,6	9,5
2	4,8	25,5	37,2	32,6
3	5,7	69,5	44,2	57,9

Una segunda aproximación para predecir las cuotas de mercado es un modelo probabilístico, sea BTL o el modelo logit. Ambos modelos evalúan la preferencia relativa de cada producto y estiman la proporción de veces que un encuestado o un conjunto de encuestados comprará el producto. En el análisis HATCO, los valores de preferencia agregados para los productos fueron 2,4; 4,8 y 5,7 para los productos 1, 2 y 3, respectivamente. Las participaciones en el mercado utilizando el modelo BTL son entonces calculadas así:

$$\text{Cuota de mercado}_{\text{producto 1}} = 2{,}4/(2{,}4 + 4{,}8 + 5{,}7) = 0{,}186 \text{ o } 18{,}6 \text{ por ciento}$$
$$\text{Cuota de mercado}_{\text{producto 2}} = 4{,}8/(2{,}4 + 4{,}8 + 5{,}7) = 0{,}372 \text{ o } 37{,}2 \text{ por ciento}$$
$$\text{Cuota de mercado}_{\text{producto 3}} = 5{,}7/(2{,}4 + 4{,}8 + 5{,}7) = 0{,}442 \text{ o } 44{,}2 \text{ por ciento}$$

Resultados similares se obtienen utilizando el modelo probabilístico logit, que se muestra también en la Tabla 7.12. Utilizando el modelo recomendado en situaciones que implican elecciones repetitivas (modelos probabilísticos), como es el caso del limpiador industrial, HATCO tiene estimaciones de cuota de mercado que indican un orden de preferencia del producto 3, después del producto 2 y finalmente del producto 1. Habría que recordar que estos resultados son agregados para la muestra entera y que las cuotas de mercado pueden diferir dentro de segmentos específicos de los encuestados.

Resumen

El análisis conjunto pone más énfasis en la capacidad del investigador o el ejecutivo de teorizar acerca del comportamiento en la elección en lugar de hacerlo en la técnica analítica. La bondad del diseño experimental y los supuestos en relación con la forma del modelo y los tipos de relaciones entre las variables son más críticos que la elección de la técnica de estimación. Así, debería de verse fundamentalmente como un análisis exploratorio, dado que muchos de sus resultados son directamente atribuibles a los supuestos básicos realizados durante el curso del diseño y la ejecución

del estudio. Además, aunque el análisis conjunto tiene sólidos fundamentos teóricos, sus aplicaciones han aumentado considerablemente sin el correspondiente desarrollo teórico. La interpretación crítica entre el modelo conceptual de toma de decisiones asumido y los elementos apropiados del análisis conjunto hace de éste un método multivariante único. El investigador debe evaluar con precisión muchas facetas del proceso de toma de decisiones para emplear el análisis conjunto. Nos hemos centrado en ofrecer una buena comprensión de los principios de los experimentos del análisis conjunto y la forma en que representan el proceso de elección del consumidor. Esta comprensión debería permitir a los investigadores evitar malas aplicaciones de esta poderosa y relativamente nueva técnica siempre que se tengan que enfrentar a entender los juicios que intervienen en una elección y las estructuras de preferencias.

Cuestiones de revisión

1. Pregunte a tres de sus compañeros de clase que evalúen las combinaciones de elección basadas en estas variables y niveles relativos a su estilo de libro de texto preferido para las clases y especifique la regla de composición que piensa usted que utilizarán. Recoja la información tanto con los métodos de perfil completo como con los de *trade-off*.

Factor	*Nivel*
Profundidad	a. Trata con gran profundidad toda la materia. b. Introduce cada tema de forma general.
Ilustraciones	a. Cada capítulo incluye dibujos humorísticos. b. Se presentan los temas de forma ilustrada. c. Cada capítulo incluye gráficos para ilustrar los temas numéricos.
Referencias	a. Las referencias generales se incluyen al final del capítulo. b. Cada capítulo incluye referencias específicas para cada tema tratado.

¿Fue difícil para los encuestados tratar los prolijos y ligeramente abstractos conceptos que se les pedía que evaluasen? ¿Cómo mejorarían las descripciones de los factores o niveles? ¿Qué método de presentación fue más fácil para los encuestados?

2. Utilizando el método de procedimiento numérico sencillo discutido en el texto o un programa informático, analice los datos del experimento de la cuestión 1.
3. Diseñe un experimento de análisis conjunto con al menos cuatro variables y dos niveles para cada variable que sea apropiado para una decisión de marketing. Al hacerlo, defina la regla de composición que va a utilizar, el diseño experimental para la creación de estímulos y el método de análisis empleado. Utilice al menos cinco encuestados para apoyar su razonamiento.
4. ¿Cuáles son los límites prácticos del análisis conjunto en términos de variables o tipos de valores para cada variable? ¿Qué tipo de problemas de elección se ajustan mejor para analizarlos con el análisis conjunto? ¿Cuáles se ajustan menos?
5. ¿Cómo aconsejaría a un investigador de mercado a la hora de elegir entre tres tipos de metodologías conjuntas? ¿Cuáles son los supuestos más importantes a considerar, junto con las ventajas y desventajas de cada metodología?

Bibliografía

1. Addelman, S. (1962), «Orthogonal Main-Effects Plans for Asymmetrical Factorial Experiments.» *Technometrics* 4: 21–46.

2. Akaah, I. (1991), «Predictive Performance of Self-Explicated, Traditional Conjoint, and Hybrid Conjoint Models under Alternative Data Collection Modes.»

Journal of the Academy of Marketing Science 19: 309–14.

3. Allenby, G. M., N. Arora, and J. L. Ginter (1995), «Incorporating Prior Knowledge into the Analysis of Conjoint Studies.» *Journal of Marketing Research* 32 (May): 152–62.
4. Alpert, M. (1971), «Definition of Determinant Attributes: A Comparison of Methods.» *Journal of Marketing Research* 8(2): 184–91.
5. Baalbaki, I. B., and N. K. Malhotra (1995), «Standardization versus Customization in International Marketing: An Investigation Using Bridging Conjoint Analysis.» *Journal of the Academy of Marketing Science* 23(3): 182–94.
6. Bretton-Clark (1988), *Conjoint Analyzer.* New York: Bretton-Clark.
7. Bretton-Clark (1988), *Conjoint Designer.* New York: Bretton-Clark.
8. Bretton-Clark (1988), *Simgraf.* New York: Bretton-Clark.
9. Bretton-Clark (1988), *Bridger.* New York: Bretton-Clark.
10. Carmone, F. J., Jr., and C. M. Schaffer (1995), «Review of Conjoint Software.» *Journal of Marketing Research* 32 (February): 113–20.
11. Carroll, J. D., and P. E. Green (1995), «Psychometric Methods in Marketing Research: Part 1, Conjoint Analysis.» *Journal of Marketing Research* 32 (November): 385–91.
12. Conner, W. S., and M. Zelen (1959), *Fractional Factorial Experimental Designs for Factors at Three Levels, Applied Math Series S4.* Washington, D.C.: National Bureau of Standards.
13. Elrod, T., J. J. Louviere, and K. S. Davey (1992), «An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models.» *Journal of Marketing Research* 29: 368–77.
14. Green, P. E. (1984), «Hybrid Models for Conjoint Analysis: An Exploratory Review.» *Journal of Marketing Research* 21 (May): 155–69.
15. Green, P. E., S. M. Goldberg, and M. Montemayor (1981), «A Hybrid Utility Estimation Model for Conjoint Analysis.» *Journal of Marketing* 45 (Winter): 33–41.
16. Green, P. E., and A. M. Kreiger (1988), «Choice Rules and Sensitivity Analysis in Conjoint Simulators.» *Journal of the Academy of Marketing Science* 16 (Spring): 114–27.
17. Green, P. E., and A. M. Kreiger (1991), «Segmenting Markets with Conjoint Analysis.» *Journal of Marketing* 55 (October): 20–31.
18. Green, P. E., A. M. Kreiger, and M. K. Agarwal (1991), «Adaptive Conjoint Analysis: Some Caveats and Suggestions.» *Journal of Marketing Research* 28 (May): 215–22.
19. Green, P. E., and V. Srinivasan (1978), «Conjoint Analysis in Consumer Research: Issues and Outlook.» *Journal of Consumer Research* 5 (September): 103–23.
20. Green, P. E., and V. Srinivasan (1990), «Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice.» *Journal of Marketing* 54(4): 3–19.
21. Green, P. E., and Y. Wind (1975), «New Way to Measure Consumers' Judgments.» *Harvard Business Review* 53 (July–August): 107–17.
22. Hahn, G. J., and S. S. Shapiro (1966), *A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments*, Report No. 66-C-165. Schenectady, N.Y.: General Electric Research and Development Center.
23. Huber, J. (1987), «Conjoint Analysis: How We Got Here and Where We Are,» In *Proceedings of the Sawtooth Conference on Perceptual Mapping, Conjoint Analysis and Computer Interviewing*, M. Metegrano, ed., Ketchum, Idaho: Sawtooth Software, pp. 2–6.
24. Huber, J., and W. Moore (1979), «A Comparison of Alternative Ways to Aggregate Individual Conjoint Analyses,» In *Proceedings of the AMA Educator's Conference*, L. Landon, ed., pp. 64–68. Chicago: American Marketing Association.
25. Huber, J., D. R. Wittink, J. A. Fielder, and R. L. Miller (1993), «The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice.» *Journal of Marketing Research* 30 (February): 105–14.
26. Huber, J., D. R. Wittink, R. M. Johnson, and R. Miller (1992), «Learning Effects in Preference Tasks: Choice-Based versus Standard Conjoint,» *Sawtooth Software Conference Proceedings*, M. Metegrano, ed. Ketchum, ID: Sawtooth Software, pp. 275–82.
27. Huber, J., and K. Zwerina (1996), «The Importance of Utility Balance in Efficient Choice Designs.» *Journal of Marketing Research* 33 (August): 307–17.
28. Intelligent Marketing Systems, Inc. (1993), *CONSURV—Conjoint Analysis Software, Version 3.0.* Edmonton, Alberta: Intelligent Marketing Systems.
29. Intelligent Marketing Systems, Inc. (1993), *NTELOGIT, Version 3.0.* Edmonton, Alberta: Intelligent Marketing Systems.
30. Jedidi, K., R. Kohli, and W. S. DeSarbo (1996), «Consideration Sets in Conjoint Analysis.» *Journal of Marketing Research* 33 (August): 364–72.
31. Johnson, R. M. (1975), «A Simple Method for Pairwise Monotone Regression.» *Psychometrika* 40 (June): 163–68.
32. Johnson, R. M. (1991), «Comment on Adaptive Conjoint Analysis: Some Caveats and Suggestions.» *Journal of Marketing Research* 28 (May): 223–25.
33. Johnson, R. M., and K. A. Olberts (1991), «Using Conjoint Analysis in Pricing Studies: Is One Price Variable Enough?» In *Advanced Research Techniques Forum Conference Proceedings*, pp. 164–73. Beaver Creek, Colo.: American Marketing Association, pp. 12–18.
34. Johnson, R. M., and B. K. Orme (1996), «How Many Questions Should You Ask in Choice-Based Conjoint Studies?» In *Advanced Research Techniques Forum*

Conference Proceedings, Beaver Creek, Colo.: American Marketing Association, pp. 42–49.
35. Kruskal, J. B. (1965), «Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data.» *Journal of the Royal Statistical Society* B27: 251–63.
36. Kuhfeld, W. F., R. D. Tobias, and M. Garrath (1994), «Efficient Experimental Designs with Marketing Research Applications.» *Journal of Marketing Research* 31 (November): 545–57.
37. Loosschilder, G. H., E. Rosbergen, M. Vriens, and D. R. Wittink (1995), «Pictorial Stimuli in Conjoint Analysis to Support Product Styling Decisions.» *Journal of the Marketing Research Society* 37(1): 17–34.
38. Louviere, J. J. (1988), *Analyzing Decision Making: Metric Conjoint Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences, vol. 67. Beverly Hills, Calif.: Sage.
39. Louviere, J. J., and G. Woodworth (1983), «Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data.» *Journal of Marketing Research* 20: 350–67.
40. Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.
41. Mahajan, V., and J. Wind (1991), *New Product Models: Practice, Shortcomings and Desired Improvements—Report No. 91-125*. Cambridge, Mass: Marketing Science Institute.
42. McFadden, D. L. (1974), «Conditional Logit Analysis of Qualitative Choice Behavior.» In *Frontiers in Econometrics*, P. Zarembka, ed., pp. 105–42. New York: Academic Press.
43. McLean, R., and V. Anderson (1984), *Applied Factorial and Fractional Designs*. New York: Marcel Dekker.
44. Oliphant, K., T. C. Eagle, J. J. Louviere, and D. Anderson (1992), «Cross-Task Comparison of Ratings-Based and Choice-Based Conjoint," In *Sawtooth Software Conference Proceedings*, M. Metegrano, ed., pp. 383–404. Ketchum, Idaho: Sawtooth Software.
45. Oppewal, H. (1995), «A Review of Conjoint Software.» *Journal of Retailing and Consumer Services* 2(1): 55–61.
46. Oppewal, H. (1995), «A Review of Choice-Based Conjoint Software: CBC and MINT.» *Journal of Retailing and Consumer Services* 2(4): 259–64.
47. Pinnell, J. (1994), «Multistage Conjoint Methods to Measure Price Sensitivity," In *Advanced Research Techniques Forum*, pp. 65–69. Beaver Creek, Colo.: American Marketing Association.
48. Research Triangle Institute (1996), *Trade-Off VR*. Research Triangle Park, N.C.: Research Triangle Institute.
49. Reibstein, D., J. E. G. Bateson, and W. Boulding (1988), «Conjoint Analysis Reliability: Empirical Findings.» *Marketing Science* 7 (Summer): 271–86.
50. SAS Institute, Inc. (1992) *SAS Technical Report R-109: Conjoint Analysis Examples*. Cary, N.C.: SAS Institute, Inc.
51. Sawtooth Software (1993), *Adoptive Conjoint Analysis*. Evanston, Ill.: Sawtooth Software.
52. Sawtooth Software (1993), *Choice-Based Conjoint*. Evanston, Ill.: Sawtooth Software.
53. Sawtooth Software (1993), *Conjoint Value Analysis*. Evanston, Ill.: Sawtooth Software.
54. Sawtooth Technologies (1997), *SENSUS, Version 2.0*. Evanston IL.: Sawtooth Technologies.
55. Schocker, A.D., and V. Srinivasan (1977), «LINMAP (Version II): A Fortran IV Computer Program for Analyzing Ordinal Preference (Dominance) Judgments Via Linear Programming Techniques for Conjoint Measurement.» *Journal of Marketing Research* 14 (February): 101–103.
56. Smith, Scott M. (1989), *PC-MDS: A Multidimensional Statistics Package*. Provo, Utah: Brigham Young University Press.
57. SPSS, Inc. (1990), *SPSS Categories*. Chicago: SPSS, Inc.
58. Srinivasan, V. (1988), «A Conjunctive-Compensatory Approach to the Self-Explication of Multiattitudinal Preference.» *Decision Sciences* 19 (Spring): 295–305.
59. Srinivasan, V., A. K. Jain, and N. Malhotra (1983), «Improving Predictive Power of Conjoint Analysis by Constrained Parameter Estimation.» *Journal of Marketing Research* 20 (November): 433–38.
60. Srinivasan, V., and C. S. Park (1997), «Surprising Robustness of the Self-Explicated Approach to Customer Preference Structure Measurement.» *Journal of Marketing Research* 34 (May): 286–91.
61. Steckel, J., W. S. DeSarbo, and V. Mahajan (1991), «On the Creation of Acceptable Conjoint Analysis Experimental Design.» *Decision Sciences* 22(2): 435–42.
62. Thurstone, L. L. (1927), «A Law of Comparative Judgment.» *Psychological Review* 34: 276–86.
63. Tumbush, J. J. (1991), «Validating Adaptive Conjoint Analysis (ACA) Versus Standard Concept Testing," In *Sawtooth Software Conference Proceedings*, M. Metegrano, ed. Ketchum, Idaho: Sawtooth Software, pp. 177–184.
64. van der Lans, I. A., and W. Heiser (1992), «Constrained Part-Worth Estimation in Conjoint Analysis Using the Self-Explicated Utility Model.» *International Journal of Research in Marketing* 9: 325– 44.
65. Veiens, M., M. Wedel, and T. Wilms (1996), «Metric Conjoint Segmentation Methods: A Monte Carlo Comparison.» *Journal of Marketing Research* 33 (February): 73–85.
66. Wittink, D. R., and P. Cattin (1989), «Commercial Use of Conjoint Analysis: An Update.» *Journal of Marketing* 53 (July): 91–96.
67. Wittink, D. R., J. Huber, P. Zandan, and R. M. Johnson (1992), «The Number of Levels Effect in Conjoint: Where Does It Come From, and Can It Be Eliminated?" In *Sawtooth Software Conference*

Proceedings, M. Metegrano, ed. Ketchum, Idaho: Sawtooth Software, pp. 355–64.

68. Wittink, D. R., L. Krishnamurthi, and J. B. Nutter (1982), «Comparing Derived Importance Weights Across Attributes.» *Journal of Consumer Research* 8 (March): 471–74.
69. Wittink, D. R., L. Krishnamurthi, and D. J. Reibstein (1990), «The Effect of Differences in the Number of Attribute Levels on Conjoint Results.» *Marketing Letters* 1(2): 113–29.
70. Wittink, D. R., M. Vriens, and W. Burhenne (1994), «Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections.» *International Journal of Research in Marketing* 11: 41–52.

Las preferencias del consumidor: estudio de su composición a través del análisis conjunto

SALVADOR RUIZ DE MAYA y JOSÉ LUIS MUNUERA ALEMÁN
Universidad de Murcia

Estudios sobre consumo, n.º 28

1. INTRODUCCIÓN

El campo del comportamiento del consumidor ha alcanzado un papel relevante en los últimos años para los investigadores de marketing. Sobre este aspecto basta comprobar la importancia que tienen las ponencias y comunicaciones relacionadas con este tema dentro de los congresos internacionales de marketing. Este trabajo trata de abordar uno de los aspectos más significativos del comportamiento del consumidor como es el «análisis de las preferencias».

El interés por el estudio de las preferencias en los individuos ha surgido por el incremento de la competencia en el mercado y la necesidad generada en la empresa de acometer decisiones comerciales más eficientes. Por un lado, el conocimiento de estas preferencias permitirá profundizar en los procesos de elección del consumidor ante alternativas diferentes de consumo y, por otro, influirán decisivamente en los procesos de diseño de los productos-servicios, incrementando la probabilidad de ser aceptados por los consumidores-objetivo.

Son muchas y variadas las justificaciones y los enfoques utilizados por los estudiosos del marketing para abordar este campo. En este estudio nos centraremos en el análisis de las preferencias desde un punto de vista conceptual, presentando una breve pero actualizada revisión sobre la literatura publicada hasta ahora. En una segunda parte presentaremos una breve explicación del análisis conjunto como uno de los instrumentos fundamentales, desarrollados hasta ahora, para medir y analizar las preferencias de los individuos. Una vez abordados los aspectos teóricos, hemos realizado un estudio empírico para analizar la composición de las preferencias a través de su medición mediante análisis conjunto. Sobre una muestra de individuos se ha llevado a cabo una experimentación con dos tipos de productos para analizar la importancia de los aspectos afectivos y cognitivos dentro de las preferencias de los consumidores. Las implicaciones de estos resultados en unas conclusiones y la bibliografía culminan este trabajo.

2. LAS PREFERENCIAS DEL CONSUMIDOR

Entre las escasas definiciones sobre el término «preferencia» localizadas (Zajonc y Markus, 1982; Holbrook y Hirschman, 1982; Holbrok y Corfman, 1984), destaca la de los profesores Zajonc y Markus, para los cuales «las preferencias son primeramente un fenómeno comportamental basado en lo afectivo. Una preferencia de x sobre y es una tendencia del organismo a aproximárse a x más a menudo y de manera más intensa que a y. Para los individuos esta aproximación se puede traducir en diferentes actuaciones como hacer comentarios favorables, comprar un producto, cambiar de religión, dar limosna, donar sangre o dar la vida».

A pesar de las indudables ventajas de esta definicion, en ella se está admitiendo el desconocimiento del procesamiento interno de los estímulos o *inputs* sensoriales en el individuo, ya sean cognitivos o afectivos, aunque se establece una relación directa e importante con la inmediata o futura actuación del individuo con relación a ese objeto.

Tras el análisis de las diferentes conceptualizaciones y estudios empíricos de las preferencias, según nuestro criterio, podemos definir la preferencia de un individuo por un objeto como una valoración positiva, relacionada o no con un proceso de elección, consecuencia de un proceso de eva-

luación originado por una actividad cognitiva, que supone un contexto en el que aparecen varias alternativas y en donde existe un tomador de decisiones, o evaluador, con criterios de valoración establecidos.

LOS ENFOQUES COGNITIVO Y AFECTIVO EN LAS PREFERENCIAS

Tradicionalmente, los estudios de marketing sobre el consumidor han considerado básicamente el aspecto racional del comportamiento de los individuos como consumidores, es decir, que éstos operan sobre la base de la evaluación de ventajas e inconvenientes de las alternativas elegidas. Es cierto que este tipo de comportamiento es el que caracteriza una gran cantidad de procesos de decisión, pero para una mejor comprensión del comportamiento de los individuos hay que tener en cuenta que éstos también se encuentran influenciados por aspectos afectivos en sus procesos de decisión, es decir, que deberemos considerar tanto el análisis coste-beneficio (inconvenientes y ventajas) de la clásica teoría de la utilidad, como las reacciones afectivas y emocionales que pueden matizar ese análisis.

El estudio de las preferencias y el comportamiento de los individuos ha sido abordado desde tres enfoques diferentes: cognitivo, afectivo y afectivo-cognitivo.

El enfoque cognitivo

Los estudios sobre el comportamiento del consumidor se han efectuado a través de la psicología cognitiva mediante la aproximación conocida como «procedimiento cognitivo consciente», basada en el principio de que los individuos consideran las consecuencias de sus acciones antes de decidir si llevar a cabo o no un determinado comportamiento (Derbaix y Abeele, 1985). El principio básico utilizado ha sido el de que si un producto A es preferido a otro B, la razón de esa mayor preferencia estriba en que el primero tiene una mayor utilidad para el individuo que está considerando esos dos productos. Este principio general ha sido el que han adoptado la microeconomía y el marketing. El profesor Zajonc (1986) especifica este principio con tres implicaciones básicas: 1) la preferencia por un objeto es función del grado de preferencia de sus componentes; 2) las preferencias permanecen estables durante el proceso de decisión; 3) si conocemos las preferencias podemos predecir la elección.

El enfoque afectivo

Este enfoque se basa en la preponderancia de los factores afectivos en la elección. Las razones aducidas para ello son las siguientes:

1. La percepción, como procedente de la preferencia, es subjetiva. Por tanto, no podemos mantener el supuesto cognitivo de que los consumidores percibirán por igual todas las dimensiones y atributos de un producto.
2. En muchos casos, la información facilitada por el sujeto sobre los atributos del producto y sobre la utilidad de cada uno de ellos podría representar más una justificación que la razón de su actuación.
3. Incluso cuando deliberadamente el consumidor intenta actuar a través de un procesamiento cognitivo consciente, en su persecución de la racionalidad podría verse influenciado por la intromisión en el desarrollo de la decisión de procesos inconscientes o automáticos, que constituyen la base de una gran parte del comportamiento ordinario.

El enfoque afectivo-cognitivo

En este enfoque globalizador (Zajonc, 1980 y 1986; Batra, 1986), la comprensión del desarrollo de las preferencias en los procesos de decisión de consumo, teniendo en cuenta tanto los aspectos cognitivos como las emocionales, requiere la consideración de diferentes situaciones de consumo (Figura 7.L.1).

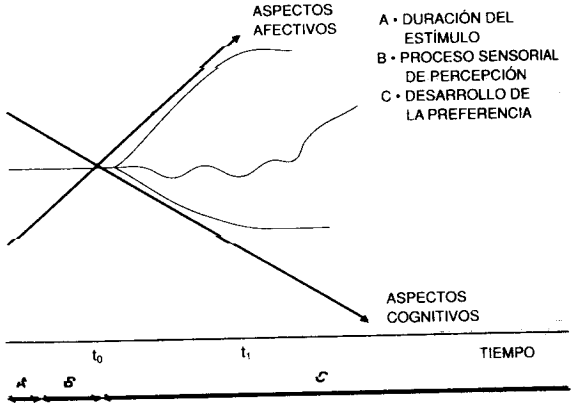


FIGURA 7.L.1. Evolución de la línea de desarrollo de la preferencia.

Fuente: Elaboración propia.

1. Situaciones de consumo en las cuales los aspectos afectivos son los responsables de la formación de la preferencia. Podemos representar el proceso de formación de las preferencias por una línea que en un espacio temporal seguiría una trayectoria como la de la Figura 7.L.1, donde se puede ver que, una vez superado el proceso sensorial, la línea alcanza un valor muy elevado en el eje de los factores afectivos, que son los que influyen en su formación, y un valor próximo a cero de los aspectos cognitivos. En un momento posterior, la formación de la preferencia se puede ver afectada por aspectos cognitivos y la curva empezará a tomar valores positivos del eje que recoge estos aspectos cognitivos, que pueden debilitar o reforzar esa preferencia.

Como señala Tsal (1985), no podemos decir que el valor de los factores cognitivos que intervienen sea cero, ya que

si bien no existen factores cognitivos en la evaluación del sabor del niño, por ejemplo, si intervienen otras variables en la formación de las preferencias como la identificación con un grupo determinado o el reforzamiento del consumo por los padres, que aunque impliquen factores afectivos principalmente, también implican factores cognitivos (como puede ser la disposición del producto en el hogar si hay un reforzamiento del consumo por los padres).

2. Situaciones de consumo en las cuales los aspectos emocionales y cognitivos conjuntamente son los responsables de la formación de las preferencias. En este caso el proceso de formación de preferencia seguiría una línea que tras superar el proceso sensorial de percepción alcanzaría valores positivos tanto en el eje de aspectos cognitivos como en el de los aspectos afectivos. La curva que representa la formación de las preferencias no tiene por qué seguir una línea recta más alejada o más cercana a cualquiera de los ejes, sino que presentará giros de acercamiento y alejamiento a los ejes, es decir, será una línea curva que se alejará del eje de los aspectos afectivos o del eje de los aspectos cognitivos dependiendo de que la información adquirida por el individuo sobre el objeto afecte a propiedades que éste evalúe en función de criterios cognitivos o afectivos respectivamente.

3. Situaciones de consumo en las cuales los aspectos cognitivos son los responsables de la formación de la preferencia. La formación y desarrolío de las preferencias sigue una línea paralela al eje de los aspectos cognitivos y con una influencia muy baja de los aspectos afectivos en su formación. No obstante, como consecuencia de la exposición repetida del individuo al objeto podría llegar un momento donde la curva se alejaría del eje de los aspectos cognitivos y tomaría valores más positivos en el eje de los aspectos afectivos.

3. EL ANÁLISIS CONJUNTO

Los diferentes modelos sobre las preferencias que se han ido desarrollando a lo largo de los años setenta y ochenta han tratado tanto de identificar y definir las variables a utilizar para su medición, como de cubrir las diferentes funciones a las que puede ir destinada esta información.

La característica más común de los diferentes métodos radica en que se basan en los modelos multiatributo, mediante los que recoge la hipótesis básica sobre las preferencias: la preferencia de un individuo por un producto depende de cuál sea la preferencia del individuo por cada uno de los atributos (Azjen y Fishbein, 1980). Los métodos de medición y análisis de las preferencias se pueden clasificar, fundamentalmente, en: 1) modelos composicionales, 2) modelos descomposicionales, 3) métodos basados en la nueva teoría de la utilidad, 4) métodos basados en la nueva teoría económica del comportamiento del consumidor, 5) análisis bayesiano de la preferencia, y 6) modelos secuenciales o no compensatorios (Lilien y Kotler, 1983; Malhotra, 1986 y 1988; Urban, Hauser y Dholakia, 1987).

Dentro de estos métodos, y más concretamente de los métodos composicionales, el análisis conjunto se presenta como uno de los más adecuados en la actualidad para el estudio de las preferencias y su composición (Green y Srinivasan, 1978 y 1990; Valera González, 1983, Múgica Grijalba, 1989a y 1989b; Vázquez Casielles, 1990). Este tipo de análisis consiste en un método de análisis multiatributo descomposicional que estima la estructura de preferencias de un consumidor, dada su evaluación global de un conjunto de alternativas. Dichas alternativas han sido preespecificadas sobre la base de una serie de atributos diferentes y de las posibles combinaciones de los mismos.

Como señalan Urban y Hauser (1980) el análisis conjunto proporciona información sobre la relación que existe entre las características del objeto y las percepciones del individuo, y sobre las características del objeto y las preferencias de ese individuo. Sin embargo, los modelos composicionales y la regresión de las preferencias se dedican a relacionar directamente las percepciones con las preferencias del individuo. Es decir, son dos formas de llegar a un mismo resultado partiendo de datos diferentes. En el caso de los modelos composicionales y de la regresión de las preferencias, al problema de una mala especificación de los atributos del análisis conjunto hay que añadir que el entrevistado no siempre es consciente de cuáles son las percepciones reales sobre las dimensiones del producto. En este caso podemos encontrarnos que atributos para los cuales especifique una baja preferencia pueden ser determinantes en su elección y viceversa.

A continuación abordamos el estudio del análisis conjunto basándonos en las diferentes etapas que implica su aplicación práctica, junto con los diferentes métodos alternativos disponibles para aplicar en cada etapa (Tabla 7.L.1).

a) *Selección del modelo de preferencia:* los modelos de preferencia utilizados en el análisis conjunto para relacionar los datos sobre las características del producto con los datos sobre la preferencia del individuo son cuatro: el modelo vectorial, el modelo de punto ideal, el modelo de la función de utilidades parciales y el modelo mixto.

El primero de los modelos mencionados, el modelo vectorial o también denominado de «criterio de composición», es similar en cuanto a su expresión matemática a los modelos composicionales de medición de preferencias; de ahí la forma lineal de la función de relación entre el nivel del atributo y la preferencia.

El modelo de punto ideal plantea que la preferencia por un estimulo está negativamente relacionada con la distancia al cuadro ponderada entre ese estímulo y el considerado ideal por el individuo; de ahí que presente una forma de parábola con abertura hacia abajo.

En el modelo de la función de utilidades parciales sólo se toman tres o cuatro valores del atributo para estimar la utilidad parcial de los mismos.

Por último, el modelo mixto es una combinación de los tres modelos anteriores utilizados en diferentes atributos.

b) *Métodos de recogida de datos:* existen varios métodos de recogida de datos: el método de los dos atributos a la vez, el método del perfil completo, el método TMT y el método interactivo con el computador.

En el método de dos atributos a la vez se pide al entrevistado rellenar tarjetas que enfrentan los niveles de dos atributos, indicando el *ranking* de preferencias entre las combinaciones de los mismos.

TABLA 7.L.1. Etapas del análisis conjunto

Etapa	*Métodos alternativos*
Selección de un modelo	Modelo vectorial, modelo de punto ideal, modelo de función de utilidades parciales y modelo mixto
Método de recogida de datos	Dos atributos al mismo tiempo, perfil completo, el procedimiento teléfono-correo-teléfono (TMT), y el método interactivo con el ordenador.
Construcción del conjunto de estímulos para el metodo de perfiles completos	Diseño maestro o total, diseño factorial fraccionado, muestra aleatoria de una distribución multivariable y diseños que cumplen el Óptimo de Pareto.
Presentación de los estímulos	Descripción verbal (con palabras o frases clave), descripción por párrafos, descripción gráfica bidimensional o tridimensional y productos físicos.
Escala de medida para la variable dependiente	Escala de clasificación, orden de clasificación, comparación pareada, comparaciones pareadas de suma constante, comparaciones pareadas graduadas y asignación de categorías.
Método de estimación	Métodos métricos (regresión múltiple), métodos no métricos (LINMAP, MONANOVA, PREFMAP y el algoritmo no métrico de Johnson) y métodos basados en probabilidades de elección *(logit y probit).*
Fiabilidad y validez de las estimaciones	Fiabilidad temporal, sobre el conjunto de atributos, sobre el conjunto de estímulos y sobre los métodos de recolección de datos. Validez cruzada.

Fuente: Elaboración propia a partir de Green y Srinivasan (1990).

El procedimiento del perfil completo utiliza juegos completos de atributos para formar estímulos que representen a los productos. A los entrevistados se les entregan un conjunto de tarjetas que describen la configuración completa de los productos o servicios que van a ser evaluados. El número total de estímulos a evaluar vendrá dado por todas las combinaciones posibles de los diferentes niveles de los atributos, aunque como veremos posteriormente en el estudio empírico, también se puede someter a evaluación sólo un subconjunto del total de combinaciones.

Otro método de recolección de datos es el procedimiento teléfono-correo-teléfono (TMT). En este caso los entrevistados son seleccionados aleatoriamente (marcando números aleatorios, por ejemplo) y en una primera entrevista telefónica se les explica el experimento y se concreta la próxima entrevista. Posteriormente, se les envía por correo el cuestionario, las tarjetas que describen los productos y determinados incentivos.

Para el método de dos factores a la vez se ha desarrollado el método interactivo con el ordenador, como forma de evitar la redundancia en la recogida de información. Consiste en un método secuencial que usa los resultados de las evaluaciones anteriores para seleccionar los siguientes pares de estímulos a evaluar, de forma que se minimice la redundancia de información solicitada al entrevistado.

c) *Construcción del conjunto de estímulos:* nos vamos a referir exclusivamente al método del perfil completo, por ser el método más utilizado y porque al requerir grandes cantidades de información se han tenido que desarrollar procedimientos que las disminuyan para realizar estimaciones (Green y Srinivasan, 1990).

El número de estímulos a evaluar por el entrevistado en el supuesto de utilización del método del perfil completo viene dado por uno de los siguientes diseños:

— *Diseño maestro o total.* Consiste en utilizar todas las combinaciones posibles de los diferentes niveles de los atributos. Se utiliza cuando se desea obtener información sobre los efectos en la preferencia de cada nivel de los atributos y de las interacciones entre los mismos.

— *Diseño factorial fraccionado.* Consiste en utilizar un subconjunto del total de estímulos a evaluar. Se utiliza cuando sólo se desea obtener información de los efectos principales de cada nivel de los atributos y de algunas interacciones.

El diseño factorial que menos información requiere del entrevistado es el diseño factorial ortogonal, que permite estimar únicamente efectos principales. En este caso, cada nivel de un atributo se cruza con cada nivel de los restantes atributos el mismo número de veces.

— Un tercer procedimiento para construir el conjunto de estímulos a evaluar cuando trabajamos con el método del perfil completo, consiste en utilizar un *conjunto de estímulos que cumplan el Óptimo de Pareto;* es decir, que ninguna opción domine a las otras en todos los atributos. Para ello deberemos tener previamente algún conocimiento sobre el

comportamiento aproximado de los atributos con respecto a la preferencia, y saber qué niveles del atributo tienen mayor preferencia y cuáles menos (Green y Srinivasan, 1990; Green y Helsen, 1988; Moore y Holbrook, 1990).

— Por último, se puede diseñar una *muestra aleatoria de una distribución multivariable.* En el caso de que todos los atributos sean continuos, se puede definir con ellos una distribución normal multivariable en función de su media, su desviación típica y las correlaciones entre los atributos. Llegados a este punto se elige un número *n* de estímulos de la distribución de una forma aleatoria.

En la Tabla 7.L.2 puede verse una comparación de los diferentes procedimientos de construcción de estímulos según diferentes criterios.

TABLA 7.L.2. Comparación entre los diferentes procedimientos de construcción de estímulos

	Diseño maestro	*Diseño factorial fraccionado ortogonal*	*Diseño factorial ortogonal con óptimo de pareto*	*Muestra aleatoria de distribución multivariable*
Cantidad de información requerida	— Elevada, todas las combinaciones posibles de los atributos.	— Muy poca, sólo sobre una pequeña fracción del diseño maestro	— Menos información que el diseño ortogonal, si lo aplicamos a un diseño de este tipo.	— Más información que el diseño ortogonal.
Información obtenida	— Sobre los efectos principales de los atributos y de las interacciones de éstos sobre las preferencias.	— Sobre los efectos principales de los atributos sobre las preferencias.	— Sobre los efectos principales de los atributos sobre las preferencias.	— Sobre los efectos principales de los atributos y sus interacciones sobre las preferencias.
Restricciones asumidas	— Ninguna	— La interacción entre los atributos no existe o es muy baja.	— La interacción entre los atributos no existe o es muy baja. — Los atributos tienen un comportamiento monótono en relación a la preferencia.	— Los atributos son continuos y siguen una distribución normal.
Otras ventajas	— Produce la máxima información y es considerado más realista al incorporar las interacciones entre los atributos.	— Disminuye el número de parámetros a estimar en relación al diseño maestro. — Produce estimaciones más acusadas.	— Evita situaciones poco realistas que pueden resultar de la combinación de atributos con funciones de utilidad parcial monótonas ante cambios en los niveles de atributos.	— Resulta adecuado cuando los atributos están muy correlacionados, o cuando las funciones de utilidad parcial de los atributos son monótonas ante cambios en los niveles de los atributos y la preferencia es ordinal.

Fuente: Elaboración propia

d) *Presentación de los estimulos:* las formas de presentar al individuo los estímulos a evaluar son las que aparecen en la Tabla 3, o combinaciones de ellas (Holbrook y Moore, 1981; Malhotra, 1988; Green y Srinivasan, 1990), donde hemos planteado una comparación en términos de ventajas e inconvenientes.

e) *Escala de medida para la variable dependiente:* centrándonos en el enfoque del perfil completo, la información proporcionada por los entrevistados con respecto a las preferencias por los productos representados en las tarjetas se puede obtener mediante dos escalas de medida:

— Mediante la ordenación de las tarjetas o perfiles de los productos de mayor a menor preferencia (escala no métrica).

— Mediante la evaluación en una escala de intervalos de menor a mayor agrado o de menor a mayor intención de compra (de 0 a 10, por ejemplo) de cada uno de los estímulos presentados (escala métrica).

f) *Metodos de estimación:* los métodos de estimación de los parámetros de la función de preferencia pueden ser de tres tipos (Varela González, 1983; Múgica Grijalba, 1989b; Green, Carmone y Smith, 1989; Green y Srinivasan, 1990):

TABLA 7.L.3. Comparación entre los diferentes procedimientos de presentación de estímulos

	Procedimiento			
	Descripción verbal	*Descripción por párrafos*	*Descripción gráfica*	*Productos físicos*
Ventajas	— Simplicidad — Eficiencia, pues permite evaluar un número elevado de estímulos.	—Descripción muy completa y realista de los estímulos. — Permite evaluar los futuros mensajes de publicidad para describir el producto.	— El individuo puede considerar grandes cantidades de información puesto que sólo tiene que visualizarla y no leerla. — Elevada homogeneidad en la percepción de los individuos ya que no tienen que imaginar cómo sería el estímulo en la realidad. — Los estímulos son más reales.	— Presentan estímulos reales al individuo. — Permiten una elevada homogeneidad en las percepciones de los individuos. — No requiere la descripción de sensaciones, sino que éstas son experimentadas por el entrevistado.
Inconvenientes	— Existe el riesgo de no describir adecuadamente los estímulos. — Resulta difícil de utilizar con atributos sensoriales, por la dificultad de describir sensaciones.	—El número de estímulos a evaluar no puede ser muy elevado debido a la cantidad de información que tiene que manejar el entrevistado, con lo que se pierde eficiencia en las estimaciones. — Resulta difícil de utilizar con atributos sensoriales, por la dificultad de describir sensaciones.	— Requiere grandes cantidades de tiempo y coste del investigador para preparar los estímulos. — El entrevistado puede percibir atributos diferentes a aquellos que interesan al investigador.	— Elevado coste económico y de tiempo para el investigador.

Fuente: Elaboración propia

— Métodos métricos, como la regresión múltiple.
— Métodos no métricos, como LINMAP, MONANOVA, PREFMAP y el algoritmo no métrico de Johnson (Johnson, 1974).
— Métodos basados en las probabilidades de elección.

De todos ellos, el más importante a nuestro entender y también el más utilizado es la regresión múltiple, que se puede aplicar tanto en aquellos casos donde la variable dependiente se mide mediante una escala de intervalo, como en aquellos donde la variable dependiente está medida con una escala de orden. Además, presenta la ventaja de proporcionar junto a las estimaciones de los parámetros las desviaciones estándar de los mismos.

Este método permite especificar diferentes modelos de preferencia para cada atributo, de forma que podremos alternar los diferentes modelos para los atributos que lo permitan (aquellos que sean continuos), y seleccionar después el modelo mixto que proporcione un mejor ajuste.

g) *Fiabilidad y validez de las estimaciones:* las aplicaciones del análisis conjunto deben someterse, al igual que la aplicación de otras técnicas, a tests de fiabilidad y validez de sus estimaciones.

La validez alude a la capacidad del método para medir el hecho que se está estudiando. Una condición necesaria, aunque no suficiente, de la validez de las medidas utilizadas es que éstas deben ser fiables. La fiabilidad se puede definir como el grado en que las medidas están libres de error y conducen a resultados consistentes (Peter, 1979; Vázquez Casielles, 1990).

Sobre las medidas de fiabilidad en análisis conjunto, algunos autores (Reibstein, Bateson y Boulding, 1988) proponen considerar cuatro tipos de fiabilidad:

TABLA 7.L.4. Variables cognitivas y afectivas en el experimento de Ratchford (1987)

Aspecto	*Características*	*Variables*
Cognitivo	— Utilidad (de uso)	— La decisión es principalmente lógica u objetiva / La decisión no es principalmente lógica u objetiva. — La decisión está basada principalmente en hechos funcionales / La decisión no está basada en hechos funcionales.
Afectivo	— Graficación del ego	— La decisión expresa la personalidad de uno mismo / La decisión no expresa la personalidad de uno mismo.
	— Aceptación social	— La decisión está basada en un conjunto de sentimientos / La decisión no está basada en un conjunto de sentimientos.
	— Placer sensorial	— La decisión está basada en el sentido de la vista, oído, tacto, olfato o gusto / La decisión no está basada en el sentido de la vista, oído, tacto, olfato o gusto

Fuente: Elaboración propia

1. *Fiabilidad temporal* para comprobar si los resultados se mantienen en el tiempo.
2. *Fiabilidad sobre el conjunto de estímulos.* Se trata de comprobar si los resultados se mantienen utilizando otro subconjunto de estímulos del diseño maestro.
3. *Fiabilidad sobre el conjunto de atributos.* Se trata de ver si varían las utilidades parciales al añadir o eliminar atributos.
4. *Fiabilidad sobre los métodos de recolección de datos.*

Por otro lado, la validez trata de contrastar si los parámetros estimados son capaces de reproducir las preferencias especificadas por los entrevistados. La mayor parte de los estudios se han centrado en contrastar la validez cruzada, en el sentido de comprobar la habilidad del modelo para predecir la posición en el orden de preferencias de un conjunto de estímulos diferentes al diseño fraccionado ortogonal, pero que son facilitados al individuo junto con los del diseño ortogonal sin que éste lo sepa.

4. ESTUDIO EMPÍRICO SOBRE PREFERENCIAS

A pesar de la abundante bibliografía sobre los aspectos cognitivos y afectivos de las preferencias, sólo dos artículos aportan datos empíricos que relacionen estos dos tipos de aspectos con uno o más productos (Ratchford, 1987; Peterson, Hoyer y Wilson, 1986).

Ratchford define dos dimensiones sobre las que posicionar diferentes productos. Una dimensión afectivo-cognitiva y otra dimensión sobre el grado de implicación de los productos, distinguiendo entre alto y bajo. Su punto de partida consiste en traducir en características materiales o inmateriales del producto las dimensiones afectivas y cognitivas. Posteriormente, a partir de esas características, se deducen variables con las que catalogar a los productos en función de esas dos dimensiones (véase Tabla 7.L.4).

En cuanto al estudio realizado por Peterson, Hoyer, Wilson, estos autores pretenden medir la influencia de la familiaridad del producto (como aspecto afectivo) en la decisión de compra y, en consecuencia, determinar su importancia en la formación de las preferencias. Una de sus principales conclusiones radica en que los individuos que realizan la primera compra de un producto tienden a basar su decisión en los niveles de familiaridad de las marcas, siempre que difieran en este atributo. Sin embargo, cuando ninguna de las marcas le es familiar al individuo el resto de atributos son los que más influyen y determinan la elección.

El objetivo de nuestro estudio consiste en cuantificar la importancia de los factores afectivos y cognitivos en la formación de las preferencias de los individuos. También hemos tratado de determinar qué variaciones se pueden producir en esa cuantificación ante la presencia de determinadas variables que caracterizan a la decisión.

Este doble objetivo se va a tratar de determinar por separado para dos muestras de individuos de diferente sexo, debido a que los productos (en nuestro caso el producto colonia) a testar son totalmente diferentes según vayan dirigidos a individuos masculinos o femeninos. Pretendemos, por tanto, obtener un posicionamiento de las preferencias para el producto en un espacio bidimensional de factores afectivos y cognitivos.

METODOLOGÍA DE LA INVESTIGACION

En la Tabla 7.L.5 aparece un esquema general de la metodología utilizada, para facilitar el seguimiento del experimento que presentamos a continuación.

Aunque en este esquema quedan suficientemente explicitadas las etapas por las que trascurre la investigación, creemos conveniente ampliar algunas de ellas. Así, en la primera etapa de la recogida de datos se pidió a los alumnos infor-

TABLA 7.L.5. Esquema general del desarrollo de la investigación

Fase		*Actividades*
Muestra		160 individuos: 98 de sexo masculino y 62 de sexo femenino.
Diseño de cuestionarios	*Determinación de atributos del producto*	Se realizaron dos reuniones de grupo con alumnos y alumnas para identificar los atributos determinantes de las colinas.
	Cuestionarios utilizados	Se han utilizado dos cuestionarios: uno para determinar los aromas a utilizar en el análisis conjunto, y otro para recoger los datos necesarios para dicho análisis.
Recogida de datos	*Etapa 1*	Se recogió información de los alumnos sobre las marcas de colonias que conocían.
	Etapa 2	Se recogió información sobre la preferencia por cinco aromas para seleccionar uno más preferido, uno indiferente y uno poco preferido.
	Etapa 3	Se recogió información sobre el orden de preferencia por 20 productos y otras características del uso del producto colonia.
Análisis de los resultados	*Importancia relativa de los atributos en las preferencias*	Mediante la aplicación del análisis conjunto se ha determinado la importancia de diferentes atributos en las preferencias.
	Contraste de los resultados	A través del análisis de la varianza se ha determinado la significatividad de la contribución de los atributos y sus interacciones a la ordenación de los productos.
	Análisis de otras variables de uso y elección	Análisis univariable de las variables de uso y elección del producto.
	Importancia relativa de los atributos por grupos	Se ha aplicado un análisis conjunto a los datos de diferentes grupos de individuos. Estos grupos se han identificado según las variables de uso y elección.

mación sobre las marcas de colonia que más habían usado y recordaban, hasta un máximo de 6. El número de diferentes marcas de colonia de hombre nombradas por los alumnos fue de 66, mientras que el número de marcas diferentes de colonia de mujer nombradas por las alumnas fue de 74.

Para la segunda etapa de la recogida de datos se seleccionaron cinco aromas de entre aquellas marcas de colonia que también se encontraban disponibles en diversos establecimientos de Murcia, y que no habían sido citadas por los alumnos o lo habían sido una sola vez.

En esta segunda etapa se realizó un análisis de frecuencia sobre los niveles de agrado por aroma. Basándonos en este análisis, tratábamos de seleccionar tres aromas de hombre y tres de mujer, que representasen a un aroma calificado de «agradable» por los estudiantes, otro calificado de «indiferente» y otro de «poco agradable». De este modo pretendíamos enfrentar posteriormente al potencial decisor a la elección de diferentes productos con diferencias significativas en el atributo aroma.

La calificación de agrado de los cinco aromas por los individuos fue muy homogénea, estando comprendida entre poco agradable e indiferente mayoritariamente, por ello decidimos tomar los tres aromas que mayor frecuencia presentasen entre las categorías de valoración «nada y poco agradable», por un lado, «indiferente», por otro y, por último, «bastante y muy agradable». Los aromas se denominaron como C, B y A, respectivamente, para cada submuestra.

En la tercera etapa de la recogida de datos se obtuvo información sobre la ordenación de preferencias de 20 productos de colonia, para aplicar posteriormente un análisis conjunto. Para la recogida de estos datos se utilizó el método de perfil completo. En primer lugar se determinaron cuáles eran los atributos relevantes para este tipo de productos sobre la base de: información publicada sobre estos productos, reuniones exploratorias con grupos de alumnos y pruebas previas del experimento. También se fijaron los niveles de los atributos a utilizar en el experimento, basándonos tanto en la información anterior como en la observación

TABLA 7.L.6. Atributos y sus niveles

Atributos	*Niveles de los atributos*	
	Para hombre	*Para mujer*
Precio	800 1.500 2.500 5.000	800 1.500 2.500 5.000
Aroma	1.- A 2.- B 3.- C	1.- A 2.- B 3.- C
Imagen	1.- ninguna, indiferente 2.- para hombre deportista, aventurero 3.- para hombre varonil, seductor 4.- para hombre romántico	1.- ninguna, indiferente 2.- para mujer liberal, independiente 3.- para mujer seductora 4.- para mujer romántica
Envase	1.- simple, de formas cuadradas 2.- estilizado, de formas redondeadas	1.- simple, de formas cuadradas 2.- estilizado, de formas redondeadas
Aplicador	1.- con atomizador 2.- sin atomizador	1.- con atomizador 2.- sin atomizador
Volumen	1.- 100 ml. 2.- 200 ml. 3.- 500 ml.	1.- 100 ml. 2.- 200 ml. 3.- 500 ml.

del mercado. Los atributos utilizados y sus niveles están recogidos en la Tabla 7.L.6.

En cuanto al número de estímulos se generaron dos diseños factoriales fraccionados ortogonales, uno para hombres y otro para mujeres, elaborando posteriormente tarjetas que representan a los productos a evaluar.

La aplicación del análisis conjunto a los datos sobre la ordenación de las tarjetas se llevó a cabo utilizando varios modelos de relación entre los datos de preferencias y los factores. En concreto, para las variables aroma, imagen, envase y aplicador se utilizó siempre el modelo discreto, por la propia definición de las variables (discretas). Sin embargo, para las variables precio y volumen se alternaron los modelos lineal, discreto y de punto ideal, puesto que las variables son continuas. En este último caso, se obtuvo el mejor ajuste del modelo, tanto para los datos de los alumnos como de las alumnas, cuando aplicamos un modelo lineal decreciente a la variable precio y un modelo lineal creciente a la variable volumen.

La fiabilidad de los datos está asegurada con un nivel de confianza del 99 por ciento tanto para la submuestra de alumnos como para la de alumnas.

Por otro lado, la validez del modelo es significativa al 98 por ciento de confianza para los alumnos y al 95 por ciento para las alumnas.

ANÁLISIS DE LAS PREFERENCIAS

A partir de los resultados obtenidos sobre la importancia de los atributos, podemos afirmar que los atributos más relevantes en la formación de preferencias sobre el producto colonia para los alumnos son el precio (con una importancia del 47,18 por ciento) y el aroma (con una importancia del 28,95 por ciento).

Sin embargo, para las alumnas los atributos más relevantes son el precio (37,14 por ciento de importancia), el aroma (24,89 por ciento de importancia) y la imagen asociada al producto (con una importancia del 22,83 por ciento). Las posiciones relativas entre los atributos por su importancia para las dos submuestras puede verse en el Gráfico 7.L.1.

Llegados a este punto, vamos a detenernos a considerar al carácter cognitivo o afectivo de los atributos con que se ha definido el producto colonia, para generar una posición en un plano bidimensional de estos factores. A partir del conjunto de reflexiones que hemos ido haciendo a lo largo de este trabajo, vamos a catalogar los atributos del producto de la siguiente manera:

— Atributos cognitivos, para los que la decisión es lógica y objetiva, basada en hechos funcionales. Estos atri-

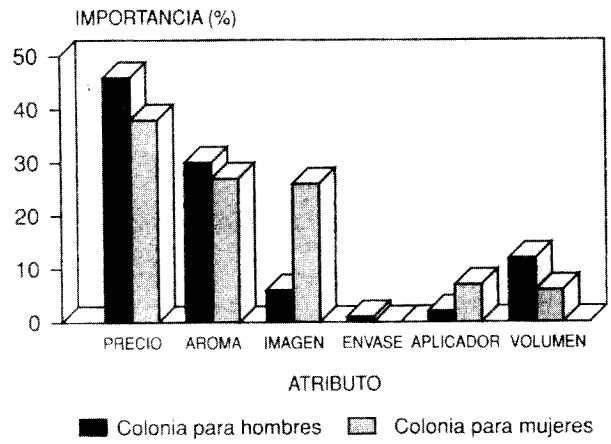


GRÁFICO 7.L.1. Comparación entre la importancia de los atributos para la submuestra de alumnos y para la submuestra de alumnas.

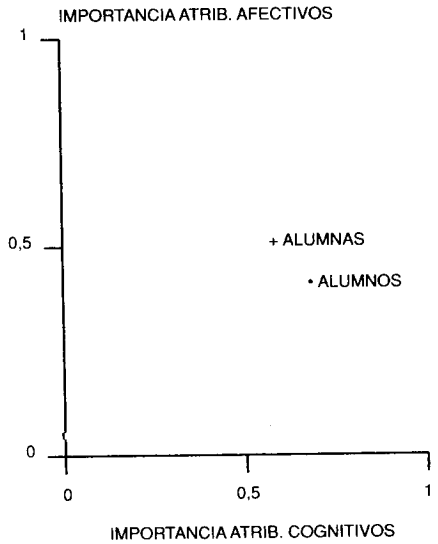


GRÁFICO 7.L.2. Representación del tipo de decisión en función de factores afectivos y cognitivos para diferentes grupos de individuos.

butos son el precio, el volumen y la existencia o no de aplicador.

— Atributos emocionales, mediante los cuales la decisión expresa la personalidad del decisor; se basa en un conjunto de sentimientos o en la agradabilidad para los sentidos. Estos atributos son la imagen asociada al producto, el aroma y la forma del envase.

Las posiciones generadas en un plano de factores afectivos y cognitivos son entonces las representadas en el Gráfico 7.L.2, a partir del cual podemos decir que en las colonias para hombre, los atributos cognitivos tienen una importancia del 63,3 por ciento en tanto que los afectivos tienen una importancia del 36,7 por ciento. Sin embargo, en las colonias para mujer los atributos cognitivos tienen una importancia del 52,16 por ciento mientras que para los efectivos es del 47,84 por ciento. Se observa claramente el predominio de los factores cognitivos sobre los afectivos para el conjunto de hombres y la parecida importancia para ambos factores en el caso de las mujeres.

CONTRASTE DE LOS RESULTADOS

El análisis conjunto realizado hasta el momento nos ha proporcionado unas estimaciones de las utilidades de los distintos atributos. No obstante, también resulta interesante conocer el nivel de significatividad de esas estimaciones. Para ello se ha realizado un análisis de la varianza (Tabla 7.L.7). Para el grupo de alumnos, entre todos los parámetros estimados, los correspondientes al precio, aroma, imagen y volumen son significativos con una probabilidad superior al 90 por ciento.

TABLA 7.L.7. Análisis de la varianza para el grupo de alumnos

Fuente de variación	*Suma de cuadrados.*	*G. L.*	*Cuadrado. medio*	*F*	*Signif.. de la f.*
Efectos principales	75,714	12	6,310	50,969	0,000
Precio	45,347	3	15,116	122,106	0,000
Aroma	16,119	2	8,059	65,103	0,000
Imagen	1,382	3	0,461	3,720	0,069
Envase	0,015	1	0,015	0,119	0,741
Aplicador	0,030	1	0,030	0,241	0,638
Volumen	5,932	2	2,516	20,326	0,001
Explicada	75,714	12	6,310	50,969	0,000
Residual	0,867	7	0,124		
TOTAL	76,581	19	4,031		

TABLA 7.L.8. Análisis de la varianza para el grupo de alumnas

Fuente de variación	*Suma de cuadrados.*	*G. L.*	*Cuadrado. medio*	*F*	*Signif.. de la f.*
Efectos principales	57,768	12	4,814	15,603	0,001
Precio	30,296	3	10,009	32,732	0,000
Aroma	21,087	2	10,543	34,173	0,000
Imagen	10,167	3	3,389	10,984	0,005
Envase	0,014	1	0,014	0,045	0,839
Aplicador	2,673	1	2,673	8,665	0,022
Volumen	1,256	2	0,628	2,036	0,201
Explicada	57,768	12	4,914	15,603	0,001
Residual	2,160	7	0,309		
TOTAL	56,928	19	3,154		

Para el grupo de alumnas los parámetros significativos con una probabilidad superior al 90 por ciento son los correspondientes al precio, aroma, imagen y la posesión de aplicador (Tabla 7.L.8).

También hemos estudiado la presencia de la multicolinealidad entre los atributos estimando las interacciones de los mismos dos a dos mediante el procedimiento ANOVA. No obstante, con los resultados de este análisis podemos seguir manteniendo nuestra afirmación de que los hombres consideran en mayor medida los atributos afectivos en la formación de sus preferencias por las colonias. Sin embargo, las mujeres parecen considerar ambos tipos de atributos casi en igual medida.

ANÁLISIS DE LAS PREFERENCIAS EN FUNCIÓN DE VARIABLES DE USO Y ELECCIÓN DEL PRODUCTO

En el cuestionario impartido a los alumnos en la tercera etapa de la investigación, además de la pregunta sobre la ordenación de las tarjetas según sus preferencias, también se incluyeron otras sobre el uso del producto y la elección del mismo como: 1) frecuencia de uso del producto, 2) actuación del individuo como consumidor solamente o como consumidor y comprador, 3) importancia del producto para mejorar la imagen del individuo, 4) familiaridad del aroma, y 5) probabilidad del compra del producto representado en la tarjeta más preferida.

La caracterización de la muestra en función de estas variables de uso y elección nos permite identificar diferentes grupos de individuos en función de las mismas. A los datos de preferencia de cada uno de esos grupos hemos aplicado un análisis conjunto para identificar la estructura de preferencias del grupo y la importancia de los atributos afectivos y cognitivos en la composición de las mismas.

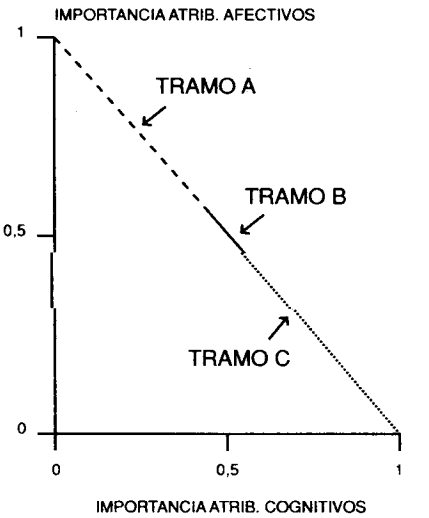


GRÁFICO 7.L.3. Posiciones en la estructura de preferencias de la importancia de los atributos afectivos y cognitivos.

La posición de los individuos en unos ejes cartesianos de atributos afectivos y cognitivos, se sitúa a lo largo de una recta que une los puntos (0,1) y (1,0) en la que podemos distinguir 3 tramos (Gráfico 7.L.3):

— Tramo A: Afectivo.
Importancia atributos cognitivos < 0,45.
Importancia atributos afectivos > 0,55.
— Tramo B: Afectivo-Cognitivo.
0,45 < Importancia atributos cognitivos < 0,55.
0,55 > Importancia atributos afectivos > 0,45.
— Tramo C: Cognitivo.
Importancia atributos cognitivos > 0,55.
Importancia atributos afectivos < 0,45.

TABLA 7.L.9. Grupos de individuos en los tramos de la recta

	Tramo A	*Tramo B*	*Tramo C*
Total Submuestras		— Total de alumnas.	— Total de alumnos.
Frecuencia de uso	— Alumnas que no usan a diario la colonia.	— Alumnas que usan a diario la colonia.	— Total de alumnos.
Consumidores y compradores	— Alumnas consumidoras.	— Alumnas consumidoras y compradoras.	— Total de alumnos.
Mejora de la imagen		— Alumnos que consideran que mejora la imagen. — Todas las alumnas.	— Alumnos que consideran que no mejora la imagen.
Experiencia en los aromas del experimento		— Alumnas sin experiencia en los aromas del experimento.	— Total de alumnos, con o sin experiencia. — Alumnas con experiencia en algún aroma del experimento.
Probabilidad de compra		— Total de alumnas.	— Total de alumnos.

En nuestro estudio podemos identificar a diferentes grupos de individuos en cada zona de la recta, según aparece en la Tabla 7.L.9.

Como se puede observar, los grupos de consumidores para los cuales los atributos afectivos son más importantes que los cognitivos en sus preferencias por este tipo de productos son:

— Mujeres que usan a diario la colonia.
— Mujeres que consumen este producto pero que no lo compran.

En el tramo C de la recta están situados todos los grupos de hombres, salvo aquellos que consideran que es un producto que contribuye a mejorar la imagen de la persona.

El grupo de mujeres para las cuales los aspectos cognitivos dominan ampliamente a los afectivos en sus preferencias por este producto son aquellas que ya conocían alguno de los aromas utilizados en el experimento.

5. CONCLUSIONES Y LIMITACIONES

En cuanto a las principales conclusiones del trabajo, cabe señalar que:

— Toda preferencia está formada por efectos cognitivos y afectivos. El mayor o menor peso de unos u otros en la formación y evolución de esa preferencia dependerá tanto del producto, como del individuo y del entorno.

— La aplicación del análisis conjunto, en su aplicación para la medición de preferencias, puede resultar más adecuada cuando se utilizan las siguientes alternativas: 1) modelo de preferencia mixto, 2) recogida de datos mediante el método del perfil completo, 3) utilización del diseño factorial fraccionado, 4) utilización de escalas de orden y de intervalos para medir el orden de preferencia de los estímulos, 5) la regresión simple como método de estimación.

— Por lo que respecta a los resultados del experimento realizado, podemos concluir que las mujeres se ven más afectadas en sus preferencias por los aspectos afectivos que los hombres. Los individuos consumidores se ven más afectados por los aspectos afectivos en sus preferencias que los individuos que son a la vez consumidores y compradores. Por último, para los siguientes grupos de individuos se produce un aumento en la importancia de los factores cognitivos con respecto al resto de consumidores:

1. Individuos que tienen experiencia con el producto que lo han probado previamente.
2. Individuos que son a la vez consumidores y compradores.
3. Individuos que tienen una mayor relación con el producto por un mayor uso.

Por otro lado, puesto que este trabajo se ha planteado como una primera aproximación a un estudio más completo sobre las preferencias, somos conscientes de las limitaciones del estudio empírico ante la ambigüedad tanto en el gusto por los aromas como en la especificación verbal de la

imagen del producto, así como la escasa representatividad de la muestra, principalmente.

Por último, en cuanto a las futuras líneas de investigación que se pretenden crear con este trabajo estarían el estudio de las preferencias familiares y de grupo, y en segundo lugar, analizar la composición de las preferencias para un mismo producto en diferentes entornos.

6. BIBLIOGRAFÍA

AIZEN, I., y FISHBEIN, M. (1980): *Understanding Attitudes and Predicting Social Behavior,* Prentice-Hall, Englewood-cliffs (New Jersey).

BATRA, R. (1986): «Affective Advertising: Role, Processes and Measurement», en R. A. Peterson, W. D. Hoyer y W. R. WITSON (eds.), *The Role of Affecting in Consumer Behaviour: Emerging Theories and Applications.* Lexington Book, Toronto, pp. 53-85.

DERBAIS, C., y ABEELE, P. V. (1985): «Consumer Inferences and Consumer Preferences: the Status of Cognition and Consciousness in Consumer Behavior Theory», *International Journal of Research in Marketing,* 2,3, pp. 157-74.

GREEN, P. E.; CARMONE Jr., F. J., y SMITH, S. M. (1989): *Multidimensional Scaling: Concepts and Applications,* Allyn and Bacon, Londres.

GREEN, P. E., y SRINIVASAN, V. (1978): «Conjoint Analysis in Consumer Research: Issues and Outlook», *Journal of Consumer Research,* 5, September, pp. 103-23.

— (1990): Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice», *Journal of Marketing,* 54, October, pp. 3-19.

HOLBROLK, M. B. (1986): «Emotion in the Consumption Experience: Toward a New Model of the Human Consumer», en R. A. Peterson, W. D. Hoyer y W. R. Wilson (eds.), *The Role of Affect in consumer Behavior: Emerging theories and Applications,* Lexington Books, Toronto, pp. 1-16.

HOLBROOK, M. B., *Consumer Perception of Merchandise and Store Quality, J. Jacoby y J. C. Olson* (eds.), y CORFMAN, K. P. (1984): «Quality and Value in the Consumption Experience: Phaedrus Rides Again», en Lexington Books, Toronto, pp. 31-57.

HOLBROOK, M. B., HIRSCHMAN, E. C. (1982): «The Experimental Aspects of Consumption: Consumer Fantasies, Feelings and Fun», *Journal of Consumer Research,* 9, September, pp. 132-40.

HOLBROOK, M. B., MOORE, W. L. (1981): «Feature Interactions in Consumer Judgments of Verbal versus Pictorial Representations», *Journal of Consumer Research* 8, June, pp. 103-13.

LILIEN, O. F., y KOTLER, P. (1983): *Marketing Decision Making: A Model Building Approach,* Harper and Row, Londres.

MALHOTRA, N. K. (1986): «An Approach to the Measurement of Consumer Preferences Using Limited Information», *Journal of Marketing Research,* 23, February, pp. 33-40.

— (1988): «A Methodology for Measuring Consumer Preferences in Developing Countries», *International Marketing Review,* 5, Autumn, pp. 52-66.

MÚGICA GRIJALBA, J. M. (1989a): «Los Modelos multiatributo en marketing: el análisis conjunto», *IP-MARK,* 324, febrero, pp. 63-71.

— (1 989b), «El análisis conjunto (y II): Alternativas, problemas y limitaciones», *IP-MARK,* 326, marzo, pp. 45-54.

O'SHAUGHINESSY, J. (1991): *Marketing competitivo,* 2.ª ed., Díaz de Santos, Madrid.

PETER, J. P. (1979): «Reliability: A Review of Psychometric Basics and Recent Marketing Practices», *Journal of Marketing Research,* 16, February, pp. 6-17.

PETERSON, R. A., y Otros (1986): «Reflection on the Role of Affect in Consumer Behavior», en W. D. Hoyer y W. R. Wilson (eds.), *The Role of Affect in Consumer Behavior: Emerging Theories and Applications,* en R. A. Perterson, Lexington Books, Toronto.

RATCHFORD, B. T. (1987): «New Insights about the FCB Grid», *Journal of Advertising Research,* 27/4, pp. 24-38.

REIBSTENIN, D. J., BATESON E. G., BOULDING, W. (1988): Conjoint Analysis Reliability: Empirical Findings», *Marketing Science,* 7, Summer, pp. 271-86.

RYAN, M. J. (1986): «Implications From the "Old" and the "New" Physics for Studying Buyer Behavior», en D. Brinberg y R. J. Lutz (eds.), *Perspectives on Metohodology in Consumer Research,* Springer, New York.

TSAL, Y. (1985): «On the Relationship Between Cognitive and Affective Processes: A Critique of Zajonc and Markus», *Journal of Consumer Research,* 12, December, pp. 358-62.

URBAN, G. L., y HAUSER, J. R. (1980): *Design and Marketing of New Product,* Prentice-Hall, Englewood Cliffs (New Jersey).

UREAN, G. L., HAUSER, J. R. y DHOLAKIA, N. (1987): *Essentials of New Product Management,* Prentice-Hall, Englewood Cliffs (New Jersey).

VARELA GONZÁLEZ, J. A. (1983): «El análisis conjunto, una técnica al servicio de la investigación comercial», *Revista de Economía y Empresa,* 11, IV, pp. 361-74.

VÁZQUEZ CASIELLES, R. (1990): «Investigación de las preferencias del consumidor mediante análisis conjunto. Importancia para el diseño de nuevos productos», *Información Comercial Española,* julio, pp. 149-63.

ZAJONC, R. B. (1980), «Feeling and Thinking: Preferences Need no Inferences», *American Psychologist,* 35, February, pp. 151-75.

— (1986): «Basic Mechanisms of Preference Formation», en R. A. Peterson, W. D. Hoyer y W. R. Wilson (eds.), *The Role of Affect in Consumer Behavior: Emerging Theories and Applications,* Lexington Books, Toronto, pp. 1-16.

ZAJONC, R. B., y MARKUS; H. (1982): «Affective and Cognitive Factors ir Preferences» *Journal of Consumer Reserch,* 9, September, pp. 123-31.

CAPÍTULO 8

Análisis de correlación canónica

OBJETIVOS DE APRENDIZAJE

- Establecer las similitudes y diferencias entre la regresión múltiple, el análisis factorial, el análisis discriminante y la correlación canónica.
- Resumir las condiciones que se deben cumplir para la aplicación del análisis de correlación canónica.
- Determinar qué mide la raíz canónica y señalar sus limitaciones.
- Establecer cuántas funciones canónicas independientes se pueden definir entre los dos conjuntos de variables originales (dependientes e independientes).
- Comparar las ventajas y desventajas de los tres métodos para interpretar la naturaleza de las funciones canónicas.
- Definir la redundancia y compararla con el R^2 de la regresión múltiple.

Hasta hace pocos años, el análisis de correlación canónica era una técnica estadística relativamente desconocida. La disponibilidad de programas de computador ha facilitado el aumento de su utilización en problemas de investigación. Es particularmente útil en situaciones donde se tienen múltiples variables dependientes como satisfacción, compra o volumen de ventas. Si las variables predictoras fueran exclusivamente categóricas, se podría emplear el análisis multivariante de la varianza. Pero, ¿qué ocurre si las variables predictoras son métricas? La correlación canónica es la respuesta, ya que permite la valoración de la relación entre variables predictoras métricas y múltiples medidas dependientes. Como se discutió en el Capítulo 1, la correlación canónica es considerada como el modelo general en el que se basan muchas otras técnicas mul-

tivariantes, dado que se pueden emplear tanto datos métricos como no métricos para variables tanto dependientes como independientes. Expresamos la forma general del análisis canónico como

$$\underset{\text{(métrica, no métrica)}}{Y_1 + Y_2 + Y_3 + \ldots + Y_n} = \underset{\text{(métrica, no métrica)}}{X_1 + X_2 + X_3 + \ldots + X_n}$$

Este capítulo introduce al analista de datos en la técnica estadística multivariante del análisis de correlación canónica. Concretamente, (1) describimos la naturaleza del análisis de correlación canónica, (2) ilustramos su aplicación y (3) discutimos sus ventajas potenciales y sus limitaciones.

¿Qué es la correlación canónica?

En el Capítulo 4 se trató el análisis de regresión múltiple, con el que se puede predecir el valor de una única variable criterio (métrica) a partir de una función lineal de un conjunto de variables predictoras (independientes). Para algunos problemas de investigación, el interés puede no centrarse en una sola variable criterio (dependiente); en su lugar, el investigador puede estar interesado en relaciones entre conjuntos de múltiples variables criterio y múltiples variables predictoras. El **análisis de correlación canónica** es un modelo estadístico multivariante que facilita el estudio de las interrelaciones entre múltiples variables criterio (dependientes) y múltiples variables predictoras (independientes) [5, 6]; es decir, mientras que la regresión múltiple predice una única variable dependiente a partir de un conjunto de múltiples variables independientes, la correlación canónica predice simultáneamente múltiples variables dependientes a partir de múltiples variables independientes.

La correlación canónica establece el menor número de restricciones sobre los tipos de datos con los que se trabaja. Dado que las otras técnicas imponen restricciones más rígidas, se acepta generalmente que la información obtenida a partir de ellas es de una mayor calidad y que se puede interpretar más fácilmente. Por esta razón, muchos investigadores consideran la correlación canónica como un último intento, para ser empleado cuando otras técnicas de mayor nivel no han funcionado correctamente. Pero en situaciones con múltiples variables dependientes e independientes, la correlación canónica es la técnica multivariante más apropiada y potente. Ha logrado la aceptación en muchos ámbitos y representa una herramienta útil para el análisis multivariante, especialmente dado que se ha extendido el interés de considerar las variables múltiples dependientes.

Ejemplo hipotético de correlación canónica

Para aclarar mejor la naturaleza de la correlación canónica, consideremos una extensión del ejemplo empleado en el Capítulo 4. Recuerde que los resultados de la encuesta de HATCO usaban el tamaño de la familia y la renta como predictores del número de tarjetas de crédito que tendría una familia. El problema implicaba el examen de la relación entre dos variables independientes y una única variable dependiente.

Suponga que HATCO estuviera interesado en un concepto más general del uso del crédito por los consumidores. Para medir este concepto, parecería lógico que HATCO considerase no solamente el número de tarjetas de crédito que tienen las familias, sino también los cargos medios mensuales en dólares que se realizan con las tarjetas de crédito. Se llegó a la conclusión de que estas dos medidas daban una mejor perspectiva sobre el uso de la tarjeta de crédito de la familia. Se remite a los lectores que tienen interés en la aproximación del uso de indicadores múltiples para representar un concepto del Capítulo 3 (Análisis factorial) y del Capítulo 11 (Modelos de ecuaciones estructurales). Ahora el problema incluye la predicción de dos medidas dependientes simultáneamente (el número de tarjetas de crédito y los cargos medios en dólares), y la regresión múltiple sólo es capaz de manejar una única variable dependiente. Se podría utilizar el análisis multivariante

TABLA 8.1. Correlación canónica del uso del crédito (número de tarjetas de crédito y porcentaje de uso) con características de los clientes (tamaño familiar y renta familiar)

Medidas de uso de crédito • Número de posesión de tarjetas de crédito • Gasto medio mensual de tarjetas de crédito en dólares		*Características de cliente* • Tamaño familiar • Ingresos familiares
Compuesto de variables dependientes	*Correlación canónica*	*Compuesto de variables independientes*
Valor teórico canónico dependiente	R_c	Valor teórico canónico independiente

de la varianza, pero únicamente si todas las variables independientes fuesen no métricas, no siendo éste el caso.

La correlación canónica es, por tanto, la única técnica disponible para examinar la relación con múltiples variables dependientes, que tienen una relación simple.

El problema de predecir el uso del crédito está reflejado en la Tabla 8.1. Las dos variables dependientes empleadas para medir el uso del crédito —número de tarjetas de crédito que tienen las familias y gastos medios mensuales en dólares realizados con las tarjetas— están representadas en el lado izquierdo. Las dos variables independientes seleccionadas para predecir el uso del crédito —el tamaño y la renta de las familias— se muestran en el lado derecho. Con el empleo del análisis de correlación canónica, HATCO puede predecir una medida compuesta del uso del crédito que incluye ambas medidas dependientes, en lugar de tener que calcular una ecuación de regresión separada para cada una de las variables dependientes. Como resultado de aplicar la correlación canónica, se obtiene una medida de la validez de la relación entre los dos conjuntos de múltiples variables (valores teóricos). Esta medida se expresa como un coeficiente de correlación canónica (R_C) entre los dos valores teóricos. Ahora el investigador tiene dos resultados de interés; los valores teóricos canónicos que representan las combinaciones lineales óptimas de las variables dependientes e independientes y la correlación canónica que representa la relación entre ellas.

El análisis de relaciones mediante la correlación canónica

El análisis de correlación canónica es el método más generalizado de la familia de las técnicas estadísticas multivariantes. Se relaciona directamente con varios métodos de dependencia. Al igual que en la regresión, el objetivo de la correlación canónica es cuantificar la validez de la relación, en este caso entre los dos conjuntos de variables (dependiente e independiente). Se asemeja al análisis factorial en la creación de compuestos de variables. También se parece al análisis discriminante en su capacidad para determinar las dimensiones independientes (igual que las funciones discriminantes) para *cada* conjunto de variables que produce la correlación máxima entre las dimensiones. De esta manera, la correlación canónica identifica la estructura óptima o la dimensionalidad de cada conjunto de variables, que maximiza la relación entre los conjuntos de variables dependientes e independientes.

El análisis de correlación canónica trata con la asociación entre los compuestos de conjuntos de variables múltiples dependientes e independientes. Por ello, desarrolla varias **funciones canónicas** que maximizan la correlación entre las **combinaciones lineales**, también conocidas como los **valores teóricos canónicos**, que son conjuntos de variables dependientes e independientes. Cada función canónica se basa realmente en la correlación entre dos valores teóricos canónicos, un valor teórico para las variables dependientes y otro para las variables independientes. Otra característica única de la correlación canónica es que se obtienen los valores teóricos de forma que se maximice su correlación. Además, la correlación canónica no acaba con la obtención de una rela-

ción simple entre los conjuntos de variables. En su lugar, se pueden conseguir varias funciones canónicas (pares de valores teóricos canónicos).

Nuestra discusión del análisis de correlación canónica se organiza alrededor del proceso de construcción del modelo descrito en el Capítulo 1. La Figura 8.1(pasos 1-3) y la Figura 8.2 (pasos 4-6) describen las pasos del análisis de correlación canónica, que incluye (1) la especificación de los objetivos de la correlación canónica, (2) el desarrollo del plan de análisis, (3) la evaluación de los supuestos básicos de la correlación canónica, (4) la estimación del modelo canónico y la valoración del ajuste global del modelo, (5) la interpretación de los valores teóricos canónicos y (6) la validación del modelo.

Primer paso: Objetivos del análisis de correlación canónica

Los datos apropiados para el análisis de correlación canónica son dos conjuntos de variables. Suponemos que a cada conjunto se le puede dar un significado teórico, al menos hasta el punto en que un conjunto pueda ser definido como las variables independientes y el otro como las variables dependientes. Una vez que se ha realizado esta distinción, la correlación canónica puede llevarse a cabo con un amplio rango de objetivos, que pueden ser algunos de los siguientes (o todos):

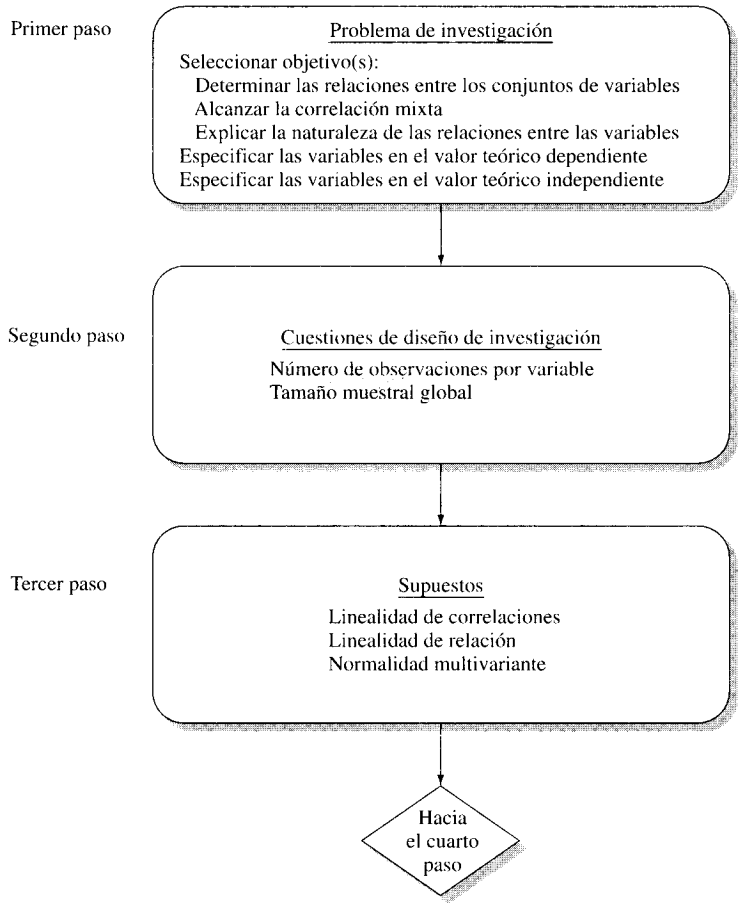


Figura 8.1. Pasos 1-3 en el diagrama de decisión de correlación canónica.

1. Determinar si dos conjuntos de variables (medidas realizadas sobre los mismos objetivos) son independientes uno de otro o, inversamente, determinar la magnitud de las relaciones que pueden existir entre los dos conjuntos.
2. Obtener un conjunto de ponderaciones para cada conjunto de variables criterio y variables predictoras, para que las combinaciones lineales de cada conjunto estén correlacionadas de forma máxima. Las funciones lineales adicionales que maximizan la restante correlación son independientes del (los) conjunto(s) anteriores de combinaciones lineales.
3. Explicar la naturaleza de cualquiera de las relaciones existentes entre los conjuntos de variables criterio y variables predictoras, generalmente midiendo la contribución relativa de cada variable a las funciones canónicas (relaciones) que son extraídas.

Segundo paso: Diseño de un análisis de correlación canónica

Al ser la forma más general del análisis multivariante, el análisis de correlación canónica comparte las herramientas básicas comunes a todas las técnicas multivariantes. Las discusiones hechas en otros capítulos (particularmente la regresión múltiple, el análisis discriminante y el análisis factorial) sobre el impacto del error de medición, los tipos de variables y sus transformaciones, son también importantes en el análisis de correlación canónica.

Frecuentemente con la correlación canónica se deben resolver cuestiones acerca del impacto del tamaño muestral (tanto pequeño como grande) y la necesidad de una cantidad suficiente de observaciones por variable. Los investigadores pueden tener la tentación de incluir muchas variables, tanto en el conjunto de variables independientes como el de dependientes, ignorando sus implicaciones en el tamaño muestral. Los tamaños muestrales que son muy pequeños, no representarán las correlaciones adecuadamente y como consecuencia esconderán cualquier relación significante que pueda existir. Los tamaños muestrales que son muy grandes, tendrán una tendencia a indicar una significación estadística en todas las instancias, incluso donde la significación práctica no está indicada. También se alienta al investigador a mantener por lo menos diez observaciones por variable para evitar el «sobreajuste» de los datos.

La clasificación de las variables como dependientes o independientes tiene poca importancia en la estimación estadística de las funciones canónicas, ya que el análisis de correlación canónica pondera ambos valores teóricos para maximizar la correlación y no establece ningún énfasis particular en alguno de los valores teóricos. Aunque dado que la técnica produce valores teóricos que maximizan la correlación entre ellos, un valor teórico en cualquier conjunto relaciona a todas las otras variables en ambos conjuntos. Con ello se permite la incorporación o la supresión de una sola variable que afecte a la solución total, particularmente el otro valor teórico. La composición de cada valor teórico, ya sea dependiente o independiente, llega a ser muy importante. El investigador, antes de aplicar el análisis de correlación canónica, debe relacionar conceptualmente los dos conjuntos de variables. De esta forma, la especificación de los valores teóricos dependientes frente a los independientes es esencial para establecer una base conceptual fuerte para las variables.

Tercer paso: Supuestos básicos de la correlación canónica

La generalidad del análisis de correlación canónica también se extiende a sus supuestos estadísticos básicos. El supuesto de linealidad afecta a dos aspectos de los resultados de la correlación canónica. Primero, el coeficiente de correlación entre cualesquiera dos variables está basado en una

relación lineal. Si la relación no es lineal, entonces se debe transformar una o ambas variables si fuera posible. Segundo, la correlación canónica es la relación lineal entre los valores teóricos. Si los valores teóricos se relacionan de una manera no lineal, la relación no será reflejada por la correlación canónica. De esta manera, aunque el análisis de correlación canónica es el método multivariante más extendido, está restringido a la identificación de relaciones lineales.

El análisis de correlación canónica puede emplear cualquier variable métrica sin que cumpla el estricto supuesto de normalidad. La normalidad es deseable porque estandariza una distribución que nos permite una mayor correlación entre las variables. Pero en un estricto sentido, el análisis de correlación canónica puede utilizar incluso variables no-normales si la forma de la distribución (p.e., altamente asimétrica) no disminuye la correlación con otras variables. Con esto se permite la utilización de datos no métricos transformados (en la forma de variables ficticias) también. Sin embargo, se requiere normalidad multivariante para los contrastes de significación de inferencia estadística de cada función canónica. Dado que los contrastes de normalidad multivariante no se están disponibles fácilmente, la línea a seguir que prevalece es asegurar que cada variable presenta normalidad univariante. De este modo, aunque estrictamente no se requiere normalidad, es altamente recomendable que se compruebe la normalidad de todas las variables y que se transformen si fuese necesario.

La homoscedasticidad, también debe ser estudiada, puesto que disminuye la correlación entre las variables. Finalmente, la multicolinealidad entre algún conjunto de variables distorsionará la capacidad de la técnica para aislar el impacto de cualquier variable única, haciendo que la interpretación sea menos fiable. Los lectores no familiarizados con estos supuestos estadísticos, los contrastes para su diagnosis o las soluciones alternativas cuando los supuestos no se cumplen, deben repasar el Capítulo 2.

Cuarto paso: Obtención de las funciones canónicas y valoración del ajuste global

El primer paso del análisis de correlación canónica es la obtención de una o más funciones canónicas (véase Figura 8.2). Cada función está formada por un par de valores teóricos, uno que representa las variables independientes y el otro que representa las variables dependientes. El número máximo de valores teóricos canónicos (funciones) que se pueden obtener a partir de los conjuntos de variables es igual al número de variables que hay en el conjunto de datos menor, ya sea dependiente o independiente. Por ejemplo, cuando un problema de investigación incluye cinco variables independientes y tres variables dependientes, el máximo número de funciones canónicas que se puede obtener es tres.

La obtención de funciones canónicas

La obtención de sucesivos valores teóricos canónicos es similar al procedimiento empleado en el análisis factorial sin rotación (véase Capítulo 3). El primer factor extraído explica la máxima cantidad de varianza en el conjunto de variables. Después se calcula el segundo factor para que explique lo más posible la varianza no explicada por el primer factor, y así sucesivamente, hasta que todos los factores han sido considerados. Por tanto, los posteriores factores se calculan a partir de los residuos o de la varianza restante de los primeros factores. El análisis de correlación canónica sigue un procedimiento similar pero centrándose en la explicación de la cantidad máxima de relación entre los dos conjuntos de variables, en lugar de en un solo conjunto. El resultado es que el primer par de valores teóricos se calcula con el fin de obtener la mayor intercorrelación posible entre los dos conjuntos de variables. El segundo par de valores teóricos canónicos es obtenido después para que represente la máxima relación entre los dos conjuntos de variables (valores teóricos) que no se ha explicado por el primer par de valores teóricos. En resumen, los sucesivos pares de valores teóricos canónicos están basados en la varianza residual, y sus respectivas correlaciones

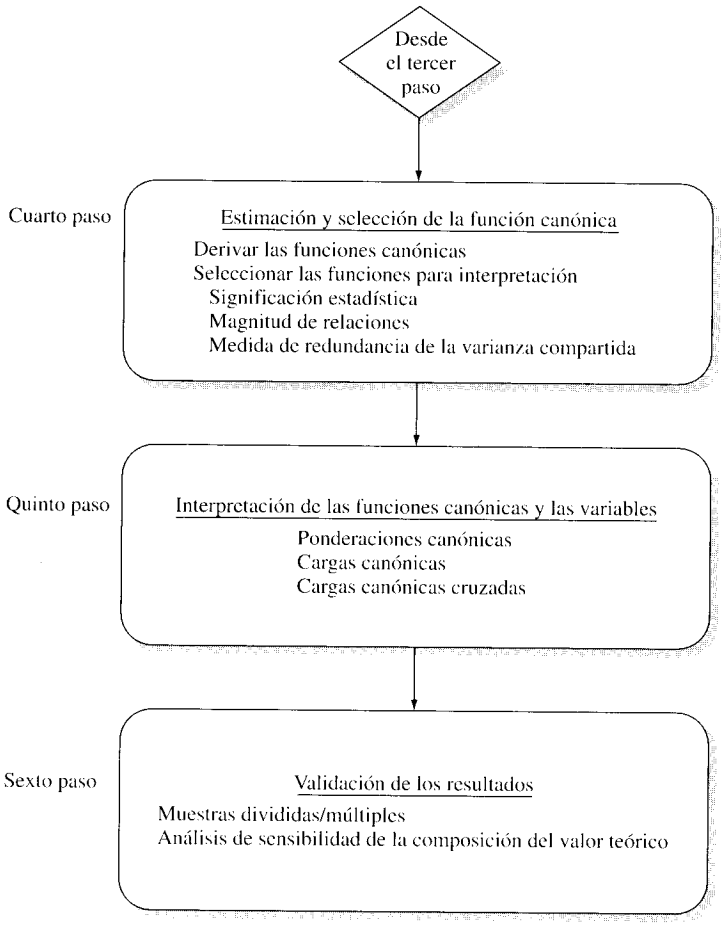


Figura 8.2. Pasos 4-6 en el diagrama de decisión de correlación canónica.

canónicas (que reflejan las interrelaciones entre los valores teóricos) disminuyen a medida que se calculan funciones adicionales; es decir, el primer par de valores teóricos canónicos refleja la mayor intercorrelación, el siguiente par la segunda correlación mayor, y así sucesivamente.

Se debe tener en cuenta un aspecto adicional acerca de la obtención de valores teóricos canónicos. Ya se ha mencionado anteriormente que los sucesivos pares de valores teóricos canónicos están basados en la varianza residual. Por tanto, cada uno de los pares de valores teóricos es **ortogonal** e independiente respecto a todos los otros valores teóricos obtenidos a partir del mismo conjunto de datos.

La validez de la relación entre los pares de valores teóricos se refleja en la correlación canónica. Cuando se eleva al cuadrado, la correlación canónica representa la cantidad de varianza de un valor teórico explicada por el otro valor teórico. A esto también se le puede definir como la cantidad de varianza compartida entre los dos valores teóricos canónicos. Las correlaciones canónicas al cuadrado son denominadas **raíces canónicas** o **autovalores**.

¿Qué funciones canónicas se deben interpretar?

Al igual que cualquier investigación que utiliza otras técnicas estadísticas, la práctica más común es analizar las funciones cuyos coeficientes de correlación canónica son estadísticamente significativos para un nivel, normalmente 0,05 o mayor. Si se consideran no significativas otras funciones independientes, éstas relaciones entre las variables no se interpretan. La interpretación de los

valores teóricos canónicos en una función significativa está basada en la premisa de que las variables en cada conjunto, que contribuyen fuertemente a las varianzas compartidas por estas funciones, son consideradas como relacionadas unas con otras.

Los autores creen que el uso de un único criterio como el nivel de significación es demasiado superficial. En lugar de esto, recomiendan que sean empleados tres criterios conjuntamente para decidir qué funciones canónicas se deben interpretar. Los tres criterios son: (1) el nivel de significación estadística de las funciones, (2) la magnitud de la correlación canónica, y (3) la medida de la redundancia para el porcentaje de varianza explicado por los dos conjuntos de datos.

Nivel de significación

El nivel de significación de una correlación canónica, que generalmente se considera como el mínimo aceptable para la interpretación, es el nivel 0,05, que (junto con el nivel 0,01) se ha llegado a convertir en el nivel más habitualmente aceptado para considerar que un coeficiente de correlación es estadísticamente significativo. Este consenso se ha dado de forma importante debido a la disponibilidad de tablas para estos niveles. Sin embargo, estos niveles no son estrictamente requeridos en todas las situaciones, y los investigadores de diferentes disciplinas, frecuentemente deben basar sus resultados en niveles de significación menores. El contraste más habitualmente utilizado, y del que normalmente disponen todos los paquetes informáticos, es el estadístico F, basado en la aproximación de Rao [3].

Además de los contrastes separados para cada función canónica, también se puede emplear un contraste multivariante de todas las raíces canónicas para evaluar la significación de dichas raíces. Muchas de las medidas existentes para valorar la significación de las funciones discriminantes, incluyendo el lambda de Wilks, la traza de Hotelling, la traza de Pillai y la mayor raíz de Roy, también están disponibles. Véase Capítulo 5 para un análisis de estas medidas.

Magnitud de las relaciones canónicas

La significación práctica de las funciones canónicas, representada por el tamaño de las correlaciones canónicas, también se tiene que tener en cuenta para decidir qué funciones interpretar. No se han establecido unas líneas básicas que consideren cuáles son tamaños aceptables para las correlaciones canónicas. En su lugar, la decisión se basa habitualmente en la contribución de los resultados para una mejor comprensión del problema de investigación que se está estudiando. Parece lógico que las líneas básicas sugeridas para las cargas factoriales significativas (véase Capítulo 3) podrían ser útiles en las correlaciones canónicas, particularmente cuando se considera que las correlaciones canónicas se refieren a la varianza explicada con los valores teóricos canónicos (combinaciones lineales), y no con las variables originales.

Medida de la redundancia de la varianza compartida

Recuerde que las correlaciones canónicas al cuadrado (raíces) proporcionan una estimación de la varianza compartida entre los valores teóricos canónicos. Aunque está es una medida sencilla y atractiva de la varianza compartida, puede llevar a algunas interpretaciones incorrectas, dado que las correlaciones canónicas al cuadrado representan la varianza compartida por las combinaciones lineales de los conjuntos de las variables criterio e independientes, pero no refleja la varianza extraída de los conjuntos de variables [1]. Por ello, se puede obtener una correlación canónica relativamente fuerte entre las dos combinaciones lineales (valores teóricos canónicos), incluso aunque estas combinaciones lineales no puedan extraer porciones significativas de varianza a partir de sus respectivos conjuntos de variables.

Dado que las correlaciones canónicas que se pueden obtener son considerablemente mayores que los coeficientes de correlación múltiple y bivariante anteriormente presentados, puede existir

la tentación de suponer que el análisis canónico ha encubierto relaciones importantes de significación teórica y práctica. Sin embargo, antes de que estas conclusiones sean corroboradas, se deben llevar a cabo posteriores análisis que incluyan otras medidas distintas a la correlación canónica para determinar la cantidad de varianza de la variable dependiente explicada o compartida por las variables independientes [7].

Para superar el sesgo y la incertidumbre propios del empleo de raíces canónicas (correlaciones canónicas al cuadrado) como una medida de la varianza compartida, se ha propuesto un **índice de redundancia** [8]. Éste es el equivalente de calcular el coeficiente de correlación múltiple al cuadrado entre el conjunto predictor total y cada una de las variables en el conjunto criterio, y después promediar estos coeficientes al cuadrado para obtener un R^2 medio. Proporciona una medida resumen de la capacidad del conjunto de las variables predictoras (consideradas como un conjunto) para explicar la variación de las variables criterio (consideradas una a una). Como tal, la medida de redundancia es perfectamente análoga al estadístico R^2 de la regresión múltiple, y su valor como índice es similar.

El índice de redundancia de Stewart-Love calcula la cantidad de varianza de un conjunto de variables que puede ser explicada por la varianza de otro conjunto. Este índice sirve como una medida de explicación de la varianza, similar al cálculo del R^2 empleado en la regresión múltiple. El R^2 representa la cantidad de varianza de la variable dependiente explicada por la función de regresión de las variables independientes . En la regresión, la varianza total de la variable dependiente es igual a 1, o al 100 por cien. Recuerde que la correlación canónica es diferente de la regresión múltiple por que no trabaja con una única variable criterio sino con un conjunto criterio que esta compuesto de varias variables, y esta combinación tiene solo una porción de la varianza total de cada variable. Por esta razón, no podemos suponer que el 100 por cien de la varianza en el conjunto criterio esté disponible para que sea explicada por el conjunto predictor. El conjunto de variables predictoras se espera que explique solamente la varianza compartida del valor teórico canónico criterio. Por esta razón, el calculo del índice de redundancia es un proceso en dos pasos. El primer paso comprende calcular la cantidad de varianza compartida del conjunto de variables criterio incluida en el valor teórico canónico dependiente. El segundo paso consiste en calcular la cantidad de varianza en el valor teórico criterio que puede ser explicada por el valor teórico canónico independiente. El paso final es calcular el índice de redundancia, que se calcula multiplicando estos dos componentes.

Paso 1: ***La cantidad de varianza compartida.*** Para calcular la cantidad de varianza compartida en el conjunto criterio incluida en el valor teórico canónico criterio, consideraremos primero como se calcula el estadístico R^2 de la regresión. El R^2 es simplemente el cuadrado del coeficiente de correlación R, que representa la correlación entre la verdadera variable dependiente y el valor predicho. En el caso canónico, estamos interesados en la correlación entre el valor teórico canónico criterio y cada una de las variables criterio. Tal información puede ser obtenida a partir de las cargas criterio (L_1), que representan la correlación entre cada variable input y su propio valor teórico (analizado en más detalle en la siguiente sección). Elevando al cuadrado cada una de las cargas criterio (L_i^2), se puede obtener una medida de la variación en cada una de las variables criterio explicada por el valor teórico canónico criterio. Para calcular la cantidad de varianza compartida explicada por el valor teórico canónico, se emplea una simple media de las cargas al cuadrado.

Paso 2: ***La cantidad de varianza explicada.*** El segundo paso del proceso de redundancia comprende el porcentaje de la varianza en el valor teórico canónico criterio que puede ser explicado por el valor teórico canónico predictor, es decir, simplemente la correlación al cuadrado entre el valor teórico canónico predictor y el valor teórico canónico criterio, que se conoce de otra forma como la correlación canónica. A la correlación canónica al cuadrado se la denomina habitualmente el R^2 canónico.

Paso 3: ***El índice de redundancia.*** El índice de redundancia se calcula multiplicando los dos componentes (varianza compartida del valor teórico por la correlación canónica al cuadrado) para obtener la cantidad de varianza compartida, que puede ser explicada por cada función canónica. Para obtener un índice de redundancia alto, se debe tener una alta correlación canónica y un alto grado

de varianza compartida explicada por el valor teórico dependiente. Una alta correlación canónica por sí sola no asegura una valiosa función canónica. Se calculan índices de redundancia tanto para el valor teórico dependiente como para el independiente, aunque en la mayoría de los casos el investigador está interesado solamente en la varianza extraída del conjunto de variables dependientes, la cual proporciona una medida mucho más realista de la capacidad predictiva de las relaciones canónicas. El investigador debe observar que mientras que la correlación canónica es la misma para los dos valores teóricos en la función canónica, el índice de redundancia probablemente variará entre los dos valores teóricos, dado que cada uno tendrá una cantidad de varianza compartida diferente.

¿Cuál es el índice mínimo de redundancia aceptable necesario para justificar la interpretación de las funciones canónicas? Al igual que con las correlaciones canónicas, no existen unas líneas básicas a seguir. El investigador debe juzgar cada función canónica según su significación teórica y práctica para el problema de investigación que se está analizando, para determinar si el índice de redundancia es suficiente para justificar la interpretación. También se ha desarrollado un contraste para la significación del índice de redundancia [2], aunque no ha sido ampliamente utilizado.

Quinto paso: Interpretación del valor teórico canónico

Si la relación canónica resulta estadísticamente significativa y las magnitudes de la raíz canónica y del índice de redundancia son aceptables, el investigador aún necesita realizar interpretaciones de los resultados. La realización de estas interpretaciones comprende el examen de las funciones canónicas para determinar la importancia relativa de cada una de las variables originales en las relaciones canónicas. Se han propuesto tres métodos: (1) ponderaciones canónicas (coeficientes estandarizados), (2) cargas canónicas (correlaciones de estructura) y (3) cargas cruzadas canónicas.

Ponderaciones canónicas

El enfoque tradicional para interpretar las funciones canónicas comprende el examen del signo y la magnitud de la ponderación canónica asociada a cada variable en su valor teórico canónico. Las variables con ponderaciones relativamente mayores contribuyen más al valor teórico y viceversa. Igualmente, las variables cuyas ponderaciones tienen signos contrarios presentan una relación inversa unas de otras, y las variables con ponderaciones del mismo signo presentan una relación directa. Sin embargo, la interpretación de la importancia o contribución relativa de una variable por su ponderación canónica está sujeta a las mismas críticas asociadas con la interpretación de los coeficientes beta en las técnicas de regresión. Por ejemplo, una ponderación pequeña puede significar o bien que su correspondiente variable es irrelevante para explicar la relación o bien que ha sido apartada de la relación debido un alto grado de multicolinealidad. Otro problema del uso de las ponderaciones canónicas es que están sujetas a una inestabilidad considerable (variabilidad) de una muestra a otra. Esta inestabilidad se da porque el procedimiento de cálculo del análisis canónico genera ponderaciones que maximizan las correlaciones canónicas para una muestra determinada de conjuntos de variables dependientes e independientes observadas [7] Este problema hace que el uso de las ponderaciones canónicas para interpretar los resultados de un análisis canónico, deba realizarse con precaución.

Cargas canónicas

El empleo de las cargas canónicas ha sustituido al uso de ponderaciones canónicas como base de interpretación, debido a las deficiencias inherentes a estas últimas. Las **cargas canónicas**, también denominadas correlaciones de estructura canónica, miden la correlación lineal simple entre una va-

riable original observada del conjunto dependiente o independiente y el valor teórico canónico del conjunto. Las cargas canónicas reflejan la varianza que la variable observada comparte con el valor teórico canónico, y puede ser interpretada como una carga factorial para valorar la contribución relativa de cada variable a cada función canónica. Se considera cada función canónica independiente de forma separada, y se calcula la correlación dentro del conjunto entre variable y valor teórico. Cuanto mayor sea el coeficiente, mayor es la importancia que tiene para calcular el valor teórico canónico. Los criterios para determinar la significación de las correlaciones de estructura canónica también son los mismos que con las cargas factoriales (véase Capítulo 3).

Las cargas canónicas, al igual que las ponderaciones, pueden estar sujetas a una importante variabilidad de una muestra a otra. Esta variabilidad sugiere que las cargas, y por tanto las relaciones asociadas a ellas, pueden ser específicas de la muestra, debido a la aleatoriedad o a factores ajenos [7]. Aunque las cargas canónicas se consideran relativamente más válidas que las ponderaciones como medios para interpretar la naturaleza de las ponderaciones canónicas, el investigador debe ser precavido cuando emplea cargas para interpretar las correlaciones canónicas, particularmente al considerar la validez externa de los resultados.

Cargas cruzadas canónicas

Se ha sugerido el cálculo de las **cargas cruzadas canónicas** como una alternativa a las cargas convencionales [4]. Este procedimiento consiste en correlacionar cada una de las variables dependientes originales observadas directamente con el valor teórico canónico independiente, y viceversa. Recuerde que las cargas convencionales correlacionan las variables originales observadas con sus respectivos valores teóricos después de que los dos valores teóricos canónicos se correlacionen de forma máxima uno con el otro. Esto puede resultar algo parecido a la regresión múltiple pero difiere en que cada variable independiente, por ejemplo, está correlacionada con el valor teórico dependiente en lugar de con una única variable dependiente. De esta manera, las cargas cruzadas proporcionan una medida más directa de las relaciones entre las variables dependientes e independientes eliminando un paso intermedio incluido en las cargas convencionales. Algunos análisis canónicos no calculan las correlaciones entre las variables y los valores teóricos. En estos casos, se considera que las ponderaciones canónicas son comparables pero no equivalentes para el propósito de nuestra discusión.

El enfoque de interpretación a emplear

Se han abordado diferentes métodos para interpretar la naturaleza de las relaciones canónicas. Sin embargo, la pregunta permanece: ¿qué método debe emplear el investigador? Dado que la mayoría de los problemas canónicos requieren un computador, el investigador debe emplear aquel método que esté disponible en los paquetes estadísticos estándar. El método preferido es el de las cargas cruzadas y está presente en muchos programas de computador, pero si las cargas cruzadas no están disponibles, el investigador está obligado, o bien a calcular las cargas cruzadas a mano, o bien a seleccionar otro método de interpretación. El enfoque de las cargas canónicas es de algún modo más válido que el uso de las ponderaciones. Por tanto, siempre que sea posible se recomienda el enfoque de las cargas como una segunda alternativa al método de cargas cruzadas canónicas.

Sexto paso: Validación y diagnosis

Al igual que cualquier otra técnica multivariante, el análisis de correlación canónica debe estar sujeto a métodos de validación que aseguren que los resultados no son solamente específicos de los datos de la muestra y que pueden ser generalizados a la población. El procedimiento más directo es crear dos submuestras de los datos (si el tamaño muestral lo permite) y llevar a cabo el análisis

en cada submuestra de forma separada. Después, los resultados se pueden comparar para buscar la igualdad de las funciones canónicas, las cargas de los valores teóricos, y demás aspectos. Si se encuentran importantes diferencias, el investigador debe considerar el realizar una investigación adicional para asegurar que los resultados finales son representativos de los valores poblacionales, y no solamente de una única muestra.

Otro enfoque consiste en evaluar la sensibilidad de los resultados a la eliminación de una variable dependiente y/o independiente. Dado que el procedimiento de correlación canónica maximiza la correlación y no optimiza la interpretabilidad, las cargas y las ponderaciones canónicas pueden variar sustancialmente si una variable es eliminada de algún valor teórico. Para asegurar la estabilidad de las cargas y de las ponderaciones canónicas el investigador debe estimar múltiples correlaciones canónicas, en donde en cada una se elimina una variable dependiente o independiente diferente.

Aunque existen pocos procedimientos de diagnosis desarrollados específicamente para el análisis de correlación canónica, el investigador debe observar los resultados teniendo en cuenta las limitaciones de la técnica. Entre las limitaciones que pueden tener un mayor impacto sobre los resultados y su interpretación están las siguientes:

1. La correlación canónica refleja la varianza compartida por las combinaciones lineales de los conjuntos de variables, no la varianza extraída de las variables.
2. Las ponderaciones canónicas obtenidas para calcular las funciones canónicas están sujetas a gran inestabilidad.
3. Las ponderaciones canónicas son obtenidas para maximizar la correlación entre las combinaciones lineales, no para la varianza extraída.
4. La interpretación de los valores teóricos canónicos puede ser difícil ya que éstos se calculan para maximizar la relación, y no existen ayudas para la interpretación como pueden ser la rotación de los valores teóricos como se vio en el análisis factorial.
5. Es difícil identificar relaciones con significado entre los subconjuntos de variables dependientes e independientes dado que aún no se han desarrollado estadísticos precisos para interpretar el análisis canónico, y debemos utilizar medidas inadecuadas como las cargas y las cargas cruzadas [7].

Sin embargo, estas limitaciones no deben desanimar a la hora de utilizar la correlación canónica. Al contrario, se las menciona para aumentar la efectividad de la correlación canónica como una herramienta de investigación.

Un ejemplo ilustrativo

Para ilustrar la aplicación de la correlación canónica, empleamos variables tomadas de la base de datos introducida en el Capítulo 1. Recuerde que los datos consisten en una serie de medidas obtenidas a partir de una muestra de 100 clientes de HATCO. Las variables incluyen las clasificaciones de HATCO sobre siete atributos (X_1 a X_7) y dos medidas que reflejan los efectos de los esfuerzos de HATCO (X_9, uso de los productos de HATCO; y X_{10}, satisfacción del cliente con HATCO).

Al igual que en capítulos anteriores, la discusión de esta aplicación de análisis de correlación canónica sigue un proceso en seis pasos ya visto anteriormente en el capítulo. En cada paso los resultados ilustran las decisiones que se deben tomar en esa etapa concreta de la investigación.

Primer paso: Objetivos del análisis de correlación canónica

Para demostrar la aplicación de la correlación canónica, empleamos las nueve variables como datos de entrada. Las clasificaciones de HATCO (X_1 a X_7) son designadas como el conjunto de múltiples variables independientes o variables predictoras. Las medidas del nivel de uso y nivel de satisfac-

ción (variables X_9 y X_{10}) se especifican como el conjunto de múltiples variables dependientes o variables criterio. El problema estadístico comprende la identificación de cualquier relación latente entre las percepciones de un cliente sobre HATCO y el nivel de uso y satisfacción del cliente.

Segundo y tercer pasos: Diseño de un análisis de correlación canónica y contrastación de los supuestos

El diseño de las variables incluye dos variables dependientes métricas y siete variables independientes métricas. La base conceptual de ambos conjuntos está bien establecida, por lo que no hay necesidad de realizar formulaciones de modelos alternativos para contrastar los diferentes conjuntos de variables. Las siete variables generan un ratio de observaciones frente a variables de 13 a 1, por lo que excede correctamente el supuesto de 10 observaciones por variable. No se considera que el tamaño muestral de 100 afecta las estimaciones del error de muestreo notablemente y por tanto no debería tener impacto sobre la significación estadística de los resultados. Finalmente, las variables dependientes e independientes fueron evaluadas en el Capítulo 2 para detectar los supuestos básicos sobre la distribución que se deben dar en el análisis multivariante y pasaron todos los tests estadísticos.

Cuarto paso: Obtención de las funciones canónicas y valoración del ajuste global

El análisis de correlación canónica está restringido a obtener dos funciones canónicas, ya que el conjunto de variables dependientes contiene solamente dos variables. Para determinar el número de funciones canónicas a incluir en el paso de interpretación, el análisis se centra en el nivel de significación estadística, la significación práctica de la correlación canónica, y los índices de redundancia para cada valor teórico.

Significación estadística y práctica

El primer contraste de significación estadística es para las correlaciones canónicas de cada una de las dos funciones canónicas. En este ejemplo, ambas correlaciones canónicas son estadísticamente significativas (véase Tabla 8.2). Junto a los tests para cada función canónica de forma separada, también se llevan a cabo contrastes multivariantes para ambas funciones simultáneamente. Los contrastes estadísticos empleados son el lambda de Wilks, el criterio de Pillai, la traza de Hotelling y la mayor raíz de Roy. La Tabla 8.2 también presenta estos tests multivariantes, los cuales indican que las funciones canónicas, consideradas conjuntamente, son estadísticamente significativas a un nivel de 0,01.

Además de la significación estadística, ambas correlaciones canónicas tienen el tamaño suficiente como para ser consideradas significativas de forma práctica. El último paso es la realización del análisis de redundancia en ambas funciones canónicas.

Análisis de redundancia

Se calcula un índice de redundancia para los valores teóricos dependiente e independiente de la primera función como se refleja en la Tabla 8.3. Como se puede observar, el índice de redundancia del valor teórico criterio es importante (0,751). Sin embargo, el valor teórico predictor tiene un índice de redundancia sustancialmente menor (0,242), aunque en este caso, dado que existe una clara explicación entre las variables dependientes e independientes, este menor valor no es inesperado o problemático. La baja redundancia del valor teórico predictor se debe a la relativamente baja va-

TABLA 8.2. Análisis de correlación canónica que relaciona el nivel de uso y satisfacción con HATCO, a las percepciones de HATCO

Medidas de ajuste global del modelo para el análisis de correlación canónica

Función canónica	*Correlación canónica*	*R^2 canónica*	*Estadístico F*	*Probabilidad*
1	0,937	0,878	30,235	0,000
2	0,510	0,260	5,391	0,000

Contraste multivariantes de significación

Estadístico	*Valor*	*Estadístico F aproximado*	*Probabilidad*
Lambda de Wilks	0,090	30,235	0,000
Traza de Pillai	1,138	17,348	0,000
Traza de Hotelling	7,535	48,441	0,000
Mrc de Roy	0,878		

rianza compartida en el valor teórico predictor (0,276), y no al R^2 canónico. Atendiendo al análisis de redundancia y a los contrastes de significación estadística, la primera función debe ser aceptada.

El análisis de redundancia para la segunda función genera unos resultados bastante diferentes (véase Tabla 8.4). Primero, el R^2 canónico es sustancialmente menor (0,260). Además, ambos conjuntos de variables presentan baja varianza compartida en la segunda función (0,145 para el valor teórico dependiente y 0,082 para el valor teórico independiente). Su combinación con la raíz canónica para obtener el índice de redundancia produce un valor de 0,038 para el valor teórico dependiente y 0,021 para el valor teórico independiente. De esta manera, mientras que la segunda función es estadísticamente significativa, tiene poca significación práctica. Con ese pequeño porcentaje, se debe cuestionar la aceptación de la función. Este es un excelente ejemplo de una función canónica estadísticamente significativa que no explica significativamente una gran parte de la varianza criterio.

TABLA 8.3. Cálculo de los índices de redundancia para la primera función canónica

Valor teórico/variables	*Carga canónica*	*Carga canónica elevada al cuadrado*	*Carga media elevada al cuadrado*	*R^2 canónico*	*Índice de redundancia*[a]
Variables dependientes					
X_9 Nivel de uso	0,913	0,834			
X_{10} Nivel de satisfacción	0,936	0,876			
Valor teórico dependiente		1,710	0,855	0,878	0,751
Variables independientes					
X_1 Velocidad de entrega	0,764	0,584			
X_2 Nivel de precios	0,061	0,004			
X_3 Flexibilidad de precios	0,624	0,389			
X_4 Imagen de fabricante	0,414	0,171			
X_5 Servicio global	0,765	0,585			
X_6 Imagen de personal de venas	0,348	0,121			
X_7 Calidad de producto	−0,278	0,077			
Valor teórico independiente		1,931	0,276	0,878	0,242

[a]El índice de redundancia se calcula como la carga media elevada al cuadrado multiplicada por el R^2 canónico.

TABLA 8.4. Análisis de la redundancia de los valores teóricos dependientes e independientes para ambas funciones canónicas

Varianza estandarizada de las variables dependientes explicado por					
	Su propio valor teórico canónico (varianza compartida)			*El valor teórico canónico opuesto (redundancia)*	
Función canónica	*Porcentaje*	*Porcentaje acumulado*	R^2 *canónico*	*Porcentaje*	*Porcentaje acumulado*
1	0,855	0,855	0,878	0,751	0,751
2	0,145	1,000	0,260	0,038	0,789
Varianza estandarizada de las variables independientes explicada por					
	Su propio valor teórico canónico (varianza compartida)			*El valor teórico canónico opuesto (redundancia)*	
Función canónica	*Porcentaje*	*Porcentaje acumulado*	R^2 *canónico*	*Porcentaje*	*Porcentaje acumulado*
1	0,276	0,276	0,878	0,242	0,242
2	0,082	0,358	0,260	0,021	0,263

El investigador que tiene interés debería revisar el Capítulo 3 con atención al debate de desarrollo de escala. De alguna manera la correlación canónica es una forma de desarrollo de escala, dado que los valores teóricos dependientes e independientes representan dimensiones de los conjuntos de la variable similares a las escalas desarrolladas con el análisis factorial. La diferencia principal es que se desarrolla estas dimensiones para maximizar la relación entre ellos, mientras que el análisis factorial maximiza la explicación (varianza compartida) del conjunto de variables.

Quinto paso: Interpretación de los valores teóricos canónicos

Una vez que se ha considerado estadísticamente significativa la relación canónica y aceptables la magnitud de la raíz canónica y el índice de redundancia, el investigador procede a realizar interpretaciones de los resultados. Aunque la segunda función podría ser considerada prácticamente no significativa, debido al bajo índice de redundancia, se incluye en la fase de interpretación por razones ilustrativas.

Estas interpretaciones comprenden el examen de las funciones canónicas para determinar la importancia relativa de cada una de las variables originales para obtener las relaciones canónicas. Los tres métodos de interpretación son (1) las ponderaciones canónicas (coeficientes estandarizados), (2) cargas canónicas (correlaciones de estructura) y (3) cargas cruzadas canónicas.

Ponderaciones canónicas

La Tabla 8.5 contiene las ponderaciones canónicas estandarizadas para cada valor teórico canónico, tanto para las variables dependientes como independientes. Como se discutió anteriormente, la magnitud de las ponderaciones representa su contribución relativa al valor teórico. Basado en el tamaño de las ponderaciones, el orden de contribución de las variables independientes al primer valor teórico es X_3, X_5, X_4, X_1, X_2, X_6 y X_7, mientras que el orden de las variables dependientes sobre el primer valor teórico es X_{10} y después X_9. Clasificaciones similares se pueden encontrar para los valores teóricos de la segunda función canónica. Dado que las ponderaciones canónicas son ge-

TABLA 8.5. Ponderaciones canónicas para las dos funciones canónicas

	Ponderaciones canónicas	
	Función 1	*Función* 2
Coeficientes estandarizadas para las variables independientes		
X_1 Velocidad de entrega	0,225	−0,965
X_2 Nivel de precios	0,103	−0,868
X_3 Flexibilidad de precios	0,569	0,160
X_4 Imagen de fabricante	0,348	−1,456
X_5 Servicio global	0,445	1,530
X_6 Imagen de personal de ventas	−0,051	0,736
X_7 Calidad de producto	0,001	0,478
Coeficientes estandarizadas para las variables dependientes		
X_9 Nivel de uso	0,501	1,330
X_{10} Nivel de satisfacción	0,580	−1,298

neralmente inestables, particularmente en casos de multicolinealidad, debido a que se calculan exclusivamente para optimizar la correlación canónica, se consideran más apropiadas las cargas canónicas y las cargas cruzadas canónicas.

Cargas canónicas

La Tabla 8.6 representa las cargas canónicas para los valores teóricos dependiente e independiente para ambas funciones canónicas. El objetivo de maximizar los valores teóricos para la correlación entre ellos tiene como resultado unos valores teóricos «optimizados» no para la interpretación, sino para la predicción. Con esto, la identificación de las relaciones es más difícil. En el primer valor teórico dependiente, ambas variables tienen cargas que exceden 0,90, reflejándose en una alta varianza compartida (0,855). Con lo que se indica un alto grado de intercorrelación entre las dos variables y sugiere que ambas medidas o bien una u otra son representativas de los efectos de los esfuerzos de HATCO.

El primer valor teórico independiente sigue un modelo bastante diferente, con cargas que varían desde 0,061 a 0,765, con una variable independiente (X_7) que incluso tiene una carga negativa, aunque es bastante pequeña y no de interés sustantivo. Las tres variables con las cargas más altas en cuanto al valor teórico independiente son X_5 (servicio global), X_1 (velocidad de entrega), y X_3 (flexibilidad de precio). Este valor teórico no corresponde a las dimensiones extraídas en el análisis factorial (véase Capítulo 3). Aún así, no deberíamos esperar una gran coincidencia entre los resultados de ambas técnicas, porque los valores teóricos en la correlación canónica se extraen solamente para maximizar los objetivos de predicción. Como tal, debería corresponder más a los resultados de otras técnicas de dependencia. Hay una correspondencia estrecha con la regresión múltiple (véase Capítulo 4). Dos de estas variables (X_3 y X_5) fueron incluidas en el análisis la regresión por etapas, en el que X_9 (una de las dos variables en el valor teórico dependiente) era la variable dependiente. Por lo tanto, la primera función canónica corresponde más estrechamente a los resultados de la regresión múltiple, con el valor teórico independiente representando el conjunto de variables que realiza la mejor predicción de las dos medidas dependientes. El investigador también debe realizar un análisis de sensibilidad del valor teórico independiente en este caso para ver si las cargas cambian cuando es eliminada una variable independiente (véase paso 6).

Los bajos valores de redundancia del segundo valor teórico están reflejados por las cargas sustancialmente menores para ambos valores teóricos en la segunda función. Por ello, la interpretabilidad más pobre reflejada en las menores cargas, unido a los bajos valores de redundancia, refuerzan la baja significación práctica de la segunda función.

Cargas cruzadas canónicas

La Tabla 8.6 también incluye las cargas cruzadas para las dos funciones canónicas. Al estudiar la primera función canónica, podemos observar que ambas variables dependientes (X_9 y X_{10}) presentan altas correlaciones con el valor teórico independiente (función 1): 0,855 y 0,877 respectivamente. Esto refleja la alta varianza compartida entre estas dos variables. Elevando al cuadrado estos términos, hallamos el porcentaje de la varianza para cada una de las variables explicadas por la función 1. Los resultados muestran que el 73 por ciento de la varianza de X_9 y el 77 por ciento de la varianza de X_{10} está explicado por la función 1. Observando las cargas cruzadas de las variables independientes, vemos que las variables X_1 y X_5 tienen altas correlaciones de aproximadamente 0,72 con el valor teórico canónico criterio. De esta información, observamos que el 52 por ciento de la varianza de cada una de estas dos variables está explicada por el valor teórico criterio (el 52 por ciento se obtiene elevando al cuadrado el coeficiente de correlación; 0,72). La correlación de X_3 (0,584) puede parecer alta, pero después de elevar al cuadrado su correlación, solamente el 34 por ciento de la variación está incluida en el valor teórico canónico.

La última cuestión de la interpretación es examinar los signos de las cargas cruzadas. Todas las variables independientes excepto X_7 (calidad del producto) tienen una relación directa positiva. Para la segunda función, dos variables predictoras (X_4 y X_6), más una variable criterio (X_{10}) son negativas. Por esto todas las relaciones son directas excepto para una relación inversa en la primera función. Por lo tanto todas las relaciones son directas excepto por una sola relación inversa para la primera función.

TABLA 8.6. Estructura canónica de las dos funciones canónicas

	Ponderaciones canónicas	
	Función 1	*Función* 2
Correlaciones entre las variables independientes y sus valores teóricos canónicos		
X_1 Velocidad de entrega	0,764	0,109
X_2 Nivel de precios	0,061	0,141
X_3 Flexibilidad de precios	0,624	0,123
X_4 Imagen de fabricante	0,414	−0,626
X_5 Servicio global	0,765	0,222
X_6 Imagen de personal de ventas	0,348	−0,199
X_7 Calidad de producto	−0,278	0,219
Correlaciones entre las variables dependientes y sus valores teóricos canónicos		
X_9 Nivel de uso	0,913	0,408
X_{10} Nivel de satisfacción	0,936	−0,352
	Ponderaciones canónicas[a]	
	Función 1	*Función* 2
Correlaciones entre las variables independientes y sus valores teóricos canónicos dependientes		
X_1 Velocidad de entrega	0,716	0,056
X_2 Nivel de precios	0,058	0,072
X_3 Flexibilidad de precios	0,584	0,063
X_4 Imagen de fabricante	0,388	−0,319
X_5 Servicio global	0,717	0,113
X_6 Imagen de personal de ventas	0,326	−0,102
X_7 Calidad de producto	−0,261	0,112
Correlaciones entre las variables dependientes y sus valores teóricos canónicos		
X_9 Nivel de uso	0,855	0,208
X_{10} Nivel de satisfacción	0,877	−0,180

[a]SAS proporciona las cargas cruzadas canónicas, puesto que SPSS no lo hace.

Sexto paso: Validación y diagnosis

El último paso debe incluir una validación de los análisis de correlación canónica por medio de uno de diferentes procedimientos. Entre los enfoques disponibles estarían (1) dividir la muestra en muestras de estimación y de validación, o (2) análisis de sensibilidad del conjunto de variables independientes. La Tabla 8.7 contiene el resultado de este análisis de sensibilidad donde se examina si las cargas canónicas son estables cuando se eliminan variables independientes individuales del análisis. Como se observa, las cargas canónicas en nuestro ejemplo son fuertemente estables y consistentes en cada uno de los tres casos donde se elimina una variable independiente (X_1, X_2 o X_7). Las correlaciones canónicas totales también permanecen estables. Pero si el lector examina las ponderaciones canónicas (no reflejadas en la tabla), existirán resultados muy diferentes dependiendo de qué variable sea eliminada. Esto refuerza el procedimiento de emplear la carga canónica y la carga cruzada con objetivos de interpretación.

Una consideración gerencial de los resultados

El análisis de correlación canónica aborda dos objetivos principales: (1) la identificación de las dimensiones entre las variables dependientes e independientes que (2) maximizan la relación entre las dimensiones. Desde una perspectiva gerencial, se proporciona al investigador una visión más detallada de la estructura de los diferentes conjuntos de variables relacionadas con una relación de dependencia. Primero, los resultados indican que solamente existe una relación simple, respaldado por la baja significación práctica de la segunda función canónica. Con el estudio de esta relación, observamos primero que las dos variables dependientes están estrechamente relacionadas y crean una dimensión claramente definida para representar los resultados de los esfuerzos de HATCO. En segundo lugar, cuando una serie de variables independientes actúan como un conjunto, se

TABLA 8.7. Análisis de sensibilidad de los resultados de la correlación canónica al eliminar una variable pendiente

		Resultados después de la eliminación de		
	Valor téorico completo	X_1	X_2	X_7
Correlación canónica (R)	0,937	0,936	0,937	0,937
Raíz canónica (R^2)	0,878	0,876	0,878	0,878
VALOR TEÓRICO INDEPENDIENTE				
Cargas canónicas				
X_1 Velocidad de entrega	0,764	omitida	0,765	0,764
X_2 Nivel de precios	0,061	0,062	omitida	0,061
X_3 Flexibilidad de precios	0,624	0,624	0,624	0,624
X_4 Imagen de fabricante	0,414	0,413	0,414	0,415
X_5 Servicio global	0,765	0,766	0,766	0,765
X_6 Imagen de personal de ventas	0,348	0,348	0,348	0,348
X_7 Calidad de producto	−0,278	−0,278	−0,278	omitida
Varianza compartida	0,276	0,225	0,322	0,309
Redundancia	0,242	0,197	0,282	0,271
VALOR TEÓRICO DEPENDIENTE				
Cargas canónicas				
X_9 Nivel de uso	0,913	0,915	0,914	0,913
X_{10} Nivel de satisfacción	0,936	0,934	0,935	0,936
Varianza compartida	0,855	0,855	0,855	0,855
Redundancia	0,750	0,749	0,750	0,750

puede predecir este dimensión de resultado adecuadamente. El valor de redundancia de 0,750 sería bastante aceptable para el R^2 en una regresión múltiple comparable. Cuando se interpreta el valor teórico independiente, observamos que tres variables, X_5 (servicio global), X_1 (velocidad de entrega) y X_3 (flexibilidad de precio) proporcionan las contribuciones sustantivas y, por lo tanto, son los predictores clave de la dimensión de resultado. Éstos deberían ser los puntos de concentración en el desarrollo de cualquier estrategia que afecte a los resultados de HATCO.

Resumen

El análisis de correlación canónica es una técnica útil y potente para explorar las relaciones entre variables dependientes e independientes múltiples. La técnica es ante todo descriptiva, aunque puede ser empleada con fines predictivos. Los resultados obtenidos a partir de un análisis canónico deben dar respuestas a cuestiones relacionadas con el número de maneras en las que se relacionan dos conjuntos de múltiples variables, la validez de las relaciones y la naturaleza de las relaciones definidas.

El análisis canónico posibilita al investigador combinar en una medida compuesta, lo que de otra forma podría ser un gran número difícil de manejar de correlaciones bivariantes entre conjuntos de variables. Es útil para identificar relaciones globales entre múltiples variables dependientes e independientes, especialmente cuando el analista tiene poco conocimiento a priori sobre las relaciones entre los conjuntos de variables. Fundamentalmente, el investigador puede aplicar el análisis de correlación canónica a un conjunto de variables, seleccionar aquellas variables (tanto dependientes como independientes) que aparecen ser significativamente relacionadas, y llevar a cabo posteriores correlaciones canónicas con las restantes variables más significativas, o realizar regresiones con estas variables.

Cuestiones de revisión

1. ¿Bajo qué circunstancias seleccionaría el análisis de correlación canónica en lugar de la regresión múltiple como una técnica estadística apropiada?
2. ¿Qué tres criterios debería emplear para decidir qué funciones canónicas deben ser interpretadas? Explique el papel de cada uno.
3. ¿Cómo interpretaría un análisis de correlación canónica?
4. ¿Cuál es la relación entre la raíz canónica, el índice de redundancia y el R^2 de la regresión múltiple?
5. ¿Cuáles son las limitaciones asociadas al análisis de correlación canónica?
6. ¿Por qué el análisis de correlación canónica se ha empleado con mucha menos frecuencia que otras técnicas multivariantes?

Bibliografía

1. Alpert, Mark I., and Robert A. Peterson (1972), «On the Interpretation of Canonical Analysis» *Journal of Marketing Research* 9 (May): 187.
2. Alpert, Mark I., Robert A. Peterson, and Warren S. Marti (1975), «Testing the Significance of Canonical Correlations.» *Proceedings, American Marketing Association* 37: 117–19.
3. Bartlett M. S. (1941), «The Statistical Significance of Canonical Correlations.» *Biometrika* 32: 29.
4. Dillon, W. R., and M. Goldstein (1984), *Multivariate Analysis: Methods and Applications.* New York: Wiley.
5. Green, P. E. (1978), *Analyzing Multivariate Data.* Hinsdale, Ill.: Holt, Rinehart, & Winston.
6. Green, P. E., and J. Douglas Carroll (1978), *Mathe-*

matical Tools for Applied Multivariate Analysis. New York: Academic Press.
7. Lambert, Z., and R. Durand (1975), «Some Precautions in Using Canonical Analysis.» *Journal of Marketing Research* 12 (November): 468–75.
8. Stewart, Douglas, and William Love (1968), «A General Canonical Correlation Index.» *Psychological Bulletin* 70: 160–63.

SECCIÓN

3

TÉCNICAS DE INTERDEPENDENCIA

Valoración general

Los métodos de dependencia que se describen en la Sección 2 proporcionan al investigador varios métodos para valorar las relaciones entre una o más variables dependientes y una serie de variables independientes. Se han abordado muchos métodos que tienen en cuenta todos los tipos de escalas (métricas y no métricas) y cantidades de variables dependientes e independientes potencialmente grandes que se podrían aplicar a series de observaciones. Pero, ¿qué ocurre si se relacionan las variables o las observaciones en una manera que no puede ser representada por las relaciones de dependencia? ¿Qué ocurre si falta la valoración de interdependencia (es decir, estructura)? Una de las capacidades básicas de los seres humanos es clasificar y calificar los objetos y la información en un esquema más simple, de tal forma que podemos caracterizar los objetos dentro de los grupos como un total en vez de tener que tratar con cada objeto individual. Este es el objetivo de los métodos desarrollados en esta sección: identificar la estructura entre una serie de variables, observaciones u objetos definidos. La identificación de la estructura ofrece no solamente simplicidad, sino también una manera de descripción e incluso descubrimiento de nuevas relaciones.

Las técnicas de interdependencia, sin embargo, se enfocan únicamente sobre la definición de estructura, con la evaluación de interdependencia sin ninguna relación de dependencia asociada. Ninguna de las técnicas de interdependencia dará una definición de la estructura para optimizar o maximizar una relación de dependencia. Es tarea del investigador utilizar primero estos métodos en la identificación de la estructura y después emplearlos donde y cuando sean apropiados. Los objetivos de las relaciones de dependencia no son "mixtos" en estos métodos de interdependencia —valoran la estructura para su propio fin.

Capítulos de la Sección 3

La Sección 3 sólo tiene dos capítulos, pero en realidad hay tres técnicas de interdependencia. La primera técnica de interdependencia, el análisis factorial (el Capítulo 3) fue abordado en la Sección 2 sobre la preparación para un análisis multivariante porque nos proporciona una herramienta para entender las relaciones

entre las variables, un conocimiento fundamental para todos nuestros análisis multivariantes.

Las cuestiones de multicolinealidad y parsimonia del modelo reflejan la estructura subyacente de las variables y el análisis factorial proporciona una manera objetiva de valorar las agrupaciones de variables y la capacidad de incorporar las variables compuestas, para reflejar estas agrupaciones de variables en otras técnicas multivariantes.

Pero no son solamente las variables las que tienen estructura. Aunque supongamos que haya independencia entre las observaciones y variables en nuestra estimación de relaciones, también sabemos que la mayoría de las poblaciones tienen subgrupos que comparten características generales. Los técnicos de marketing buscan mercados objetivos de grupos diferenciados de consumidores homogéneos, los analistas estratégicos buscan grupos de empresas similares para identificar elementos de estrategia común y los analistas financieros buscan acciones con fundamentos similares para crear carteras de acciones. Éstas, junto con muchas otras situaciones, requieren técnicas que encuentran estos grupos de objetos similares basados en una serie de características.

Este es el objetivo del Capítulo 9, el análisis cluster. El análisis cluster se ajusta idealmente a la definición de grupos de objetos con homogeneidad máxima dentro de los grupos, a la vez que mantiene una heterogeneidad máxima entre los grupos —con la determinación de los grupos más similares que son también los más diferentes el uno del otro. Tal y como mostraremos, el análisis cluster tiene una tradición rica de aplicación en casi cada área de investigación. Pero su capacidad de definir grupos de objetos similares está contrarrestada por su naturaleza bastante subjetiva y el papel instrumental llevado a cabo por el juicio del investigador en varias decisiones claves. Esto no reduce la utilidad de la técnica, pero sí supone una carga mayor para el investigador a la hora de comprender completamente la técnica y el impacto de ciertas decisiones sobre la solución cluster definitiva.

Pero, ¿qué ocurre si solamente sabemos cómo son los objetos de similares y no sabemos la fuente de esta similitud o cómo agrupar mejor los objetos? Se aborda esta situación en el Capítulo 10, el análisis multidimensional. El análisis multidimensional es una técnica que comienza como un análisis univariante —una medida única de similitud entre objetos— y deduce la dimensionalidad de las similitudes entre los objetos. Intenta responder a la siguiente cuestión básica: ¿es posible agrupar los objetos en uno, dos, tres o n espacios dimensionales en una manera en que se representen adecuadamente las similitudes entre los objetos por su proximidad? Como tal, el análisis multidimensional es una forma de análisis decomposicional, un poco como el análisis conjunto (véase Capítulo 7), pero en este caso no existen características conocidas de los objetos, solamente sus similitudes. Una forma especial del análisis multidimensional es el análisis de correspondencias, que analiza una forma clara de datos —las variables categóricas de tabulación cruzada. A partir de estos datos, el análisis de correspondencias puede dibujar las relaciones entre las filas y las columnas (por ejemplo, productos y atributos) en una perspectiva dimensional en la que la proximidad representa la similitud.

El análisis cluster, el análisis factorial y el análisis multidimensional proporcionan al investigador métodos que pueden poner "orden" a los datos en la forma de estructura entre las observaciones o variables. De esta manera, el investigador puede comprender mejor las estructuras básicas de los datos, no solamente facilitando la descripción de los datos, sino también proporcionando una base para un análisis más refinado de las relaciones de dependencia.

CAPÍTULO 9

Análisis cluster

OBJETIVOS DE APRENDIZAJE

- Definir las cuestiones apropiadas a investigar para ser tratadas por el análisis cluster.
- Entender cómo se mide la similitud entre objetos.
- Distinguir entre las diferentes medidas de distancia.
- Diferenciar entre los algoritmos de cluster y sus aplicaciones adecuadas.
- Entender las diferencias entre las técnicas cluster jerárquicas y no jerárquicas.
- Entender cómo se selecciona el número de conglomerados a formar.
- Seguir las líneas principales de validación del conglomerado.
- Construir los perfiles de los conglomerados derivados y evaluar su significación práctica.
- Exponer las limitaciones del análisis cluster.

Los académicos y los investigadores de mercado se encuentran a menudo con situaciones cuya mejor forma de resolverlas es definiendo grupos de objetos homogéneos, tanto si son individuos como si son empresas, productos o incluso comportamientos. Las opciones estratégicas basadas en los grupos identificados en la población, como la segmentación y los objetivos de márketing no serían posibles sin una metodología objetiva. Esta misma necesidad se encuentra en otras áreas, que van desde las ciencias naturales (por ejemplo, creación de una taxonomía biológica para la clasificación de varios grupos de animales —insectos frente a mamíferos frente a reptiles) a las ciencias sociales (por ejemplo, análisis de varios tipos de perfiles psiquiátricos). En todos estos casos, el investigador está buscando una estructura «natural» entre las observaciones basada en un perfil multivariante.

La técnica más utilizada para este fin es el análisis cluster. El análisis cluster agrupa a los individuos y a los objetos en conglomerados, de tal forma que los objetos del mismo conglomera-

do son más parecidos entre sí que a los objetos de otros conglomerados. Lo que se intenta es maximizar la homogeneidad de los objetos dentro de los conglomerados mientras que a la vez se maximiza la heterogeneidad entre los agregados. Este capítulo explica la naturaleza y el propósito del análisis cluster y guía al investigador en la selección y uso de los diversos enfoques del mismo.

¿Qué es el análisis *cluster*?

El análisis *cluster* es la denominación de un grupo de técnicas multivariantes cuyo principal propósito es agrupar objetos basándose en las características que poseen. El análisis *cluster* clasifica **objetos** (es decir, encuestados, productos u otras entidades) de tal forma que cada objeto es muy parecido a los que hay en el conglomerado con respecto a algún criterio de selección predeterminado. Los conglomerados de objetos resultantes deberían mostrar un alto grado de homogeneidad interna (dentro del conglomerado) y un alto grado de heterogeneidad externa (entre conglomerados). Por tanto, si la clasificación es acertada, los objetos dentro de los conglomerados estarán muy próximos cuando se representen gráficamente, y los diferentes grupos estarán muy alejados.

En el análisis *cluster*, de nuevo el concepto de valor teórico es central, pero en una forma muy diferente del resto de las técnicas multivariantes. El **valor teórico del análisis cluster** es el conjunto de variables que representan las características utilizadas para comparar objetos en el análisis *cluster*. Dado que el valor teórico del análisis *cluster* incluye sólo las variables utilizadas para comparar objetos, determina el «carácter» de los objetos. El análisis *cluster* es la única técnica multivariante que no estima el valor teórico empíricamente sino que utiliza el valor teórico especificado por el investigador. El objetivo del análisis *cluster* es la comparación de objetos basándose en el valor teórico, no en la estimación del valor teórico en sí misma. Esto hace crucial la definición que dé el investigador al valor teórico para el análisis *cluster*.

El análisis *cluster* se ha denominado como análisis Q, construcción de tipología, análisis de clasificación y taxonomía numérica. Esta variedad de nombres se debe en parte al uso de los métodos de agrupación en disciplinas tan diversas como psicología, biología, sociología, economía, ingeniería y negocios. Aunque los nombres difieren entre disciplinas, todos los métodos tienen una dimensión común: clasificación de acuerdo a una relación natural [1, 2, 3, 6, 12, 16]. Esta dimensión común representa la esencia de todas las aproximaciones del análisis *cluster*. Como tal, el valor fundamental del análisis *cluster* descansa en la clasificación de los datos, tal y como sugiere la agrupación «natural» de los datos en sí misma. El análisis *cluster* es comparable al análisis factorial (véase Capítulo 3) en su objetivo de evaluar la estructura. Pero el análisis *cluster* difiere del análisis factorial en que el análisis *cluster* agrupa objetos, mientras que el análisis factorial se centra principalmente en la agrupación de variables.

El análisis *cluster* es una herramienta de análisis útil para diferentes situaciones. Por ejemplo, un investigador que haya recogido datos mediante un cuestionario puede enfrentarse a un número elevado de observaciones que no tienen sentido a menos que se clasifiquen en grupos manejables. El análisis *cluster* puede llevar a cabo objetivamente este procedimiento de reducción de datos mediante la reducción de la información de una población completa o una muestra a información sobre subgrupos pequeños y específicos. Por ejemplo, si podemos entender las actitudes de una población mediante la identificación de los principales grupos de la población, entonces debemos reducir los datos de la población completa a perfiles de ciertos grupos. De esta forma, el investigador tiene una descripción más concisa y comprensible de las observaciones, con una pérdida mínima de información.

El análisis *cluster* es muy útil cuando un investigador desea desarrollar las hipótesis concernientes a la naturaleza de los datos o para examinar las hipótesis previamente establecidas. Por ejemplo, un investigador puede creer que las actitudes hacia el consumo de refrescos normales frente a «lights» podrían utilizarse para separar a los consumidores de refrescos en segmentos lógicos o grupos. El análisis *cluster* puede clasificar consumidores de refrescos por sus actitudes hacia los

refrescos normales frente a los «light», y los conglomerados resultantes, si los hay, pueden perfilarse mediante diferencias y similitudes demográficas.

Estos ejemplos son sólo una reducida fracción de los tipos de aplicaciones del análisis *cluster*. Desde la derivación de taxonomías en biología para la agrupación de todos los organismos vivientes a clasificaciones psicológicas basadas en la personalidad y otros rasgos personales, pasando por los análisis de segmentación de los mercados, el análisis *cluster* ha tenido siempre una fuerte tradición en la agrupación de individuos. Esta tradición se ha extendido a la clasificación de objetos, incluyendo la estructura de mercado, análisis de similitudes y diferencias entre productos nuevos y evaluación del rendimiento de empresas para identificar agrupaciones basadas en las estrategias de las empresas u orientaciones estratégicas. El resultado ha sido una profusión de aplicaciones en casi todas las áreas de investigación, creando no sólo una riqueza de conocimiento en el uso del análisis de conglomerados sino también la necesidad de una mejor comprensión de la técnica para minimizar su mala utilización.

Sin embargo, junto con los beneficios del análisis *cluster* existen algunos inconvenientes. El análisis *cluster* puede caracterizarse como descriptivo, ateórico y no inferencial. El análisis *cluster* no tiene bases estadísticas sobre las cuales deducir inferencias estadísticas para una población a partir de una muestra, y se utiliza fundamentalmente como una técnica de exploratoria. Las soluciones no son únicas, en la medida en que la pertenencia al conglomerado para cualquier número de soluciones depende de muchos elementos del procedimiento y se pueden obtener muchas soluciones diferentes variando uno o más de estos elementos. Además, el análisis *cluster* siempre creará conglomerados, a pesar de la existencia de una «auténtica» estructura en los datos. Finalmente, la solución *cluster* es totalmente dependiente de las variables utilizadas como base para la medida de similitud. La adición o destrucción de variables relevantes puede tener un impacto substancial sobre la solución resultante. Por tanto, el investigador debe tener particular cuidado en evaluar el impacto de cada decisión implicada en el desarrollo de un análisis *cluster*.

¿Cómo funciona el análisis *cluster*?

La naturaleza del análisis *cluster* puede ilustrarse mediante un simple ejemplo bivariante. Suponga que un investigador de mercado desea determinar los segmentos del mercado en una comunidad reducida basándose en sus pautas de lealtad a marcas y tiendas. Se selecciona una reducida muestra de siete encuestados como contrastación de prueba de cómo se aplica el análisis *cluster*. Se miden dos medidas de lealtad —V_1 (lealtad a la tienda) y V_2 (lealtad a la marca)— para cada encuestado en una escala de 0 a 10. Los valores de cada uno de los siete encuestados se muestran en la Figura 9.1, junto con un diagrama de dispersión representando cada observación en dos variables.

El objetivo principal del análisis *cluster* es definir la estructura de los datos colocando las observaciones más parecidas en grupos. Pero para llevar a cabo esta tarea, debemos tratar tres cuestiones básicas. En primer lugar, ¿cómo medimos la similitud? Necesitamos un método de observaciones simultáneamente comparadas sobre dos variables de aglomeración (V_1 y V_2). Son posibles varios métodos, incluyendo la correlación entre objetos, una medida de asociación utilizada en otras técnicas multivariantes o quizá midiendo su proximidad en un espacio bidimensional de tal forma que la distancia entre las observaciones indica similitud. En segundo lugar, ¿cómo formamos los conglomerados? No importa cómo se mida la similitud, el procedimiento debe agrupar aquellas observaciones que son más similares dentro de un conglomerado. Este procedimiento debe determinar la pertenencia al grupo de cada observación. En tercer lugar, ¿cuántos grupos formamos?

Puede utilizarse cualquier número de «reglas», pero la tarea fundamental es evaluar la similitud «media» dentro de los conglomerados, de tal forma que a medida que la media aumenta, el conglomerado se hace menos similar. El investigador se enfrenta a continuación a un *trade-off*: pocos conglomerados frente a menos homogeneidad. Una estructura simple, al tender hacia la parsimo-

Valores de datos

Variable de cluster	Encuestados						
	A	B	C	D	E	F	G
V_1	3	4	4	2	6	7	6
V_2	2	5	7	7	6	7	4

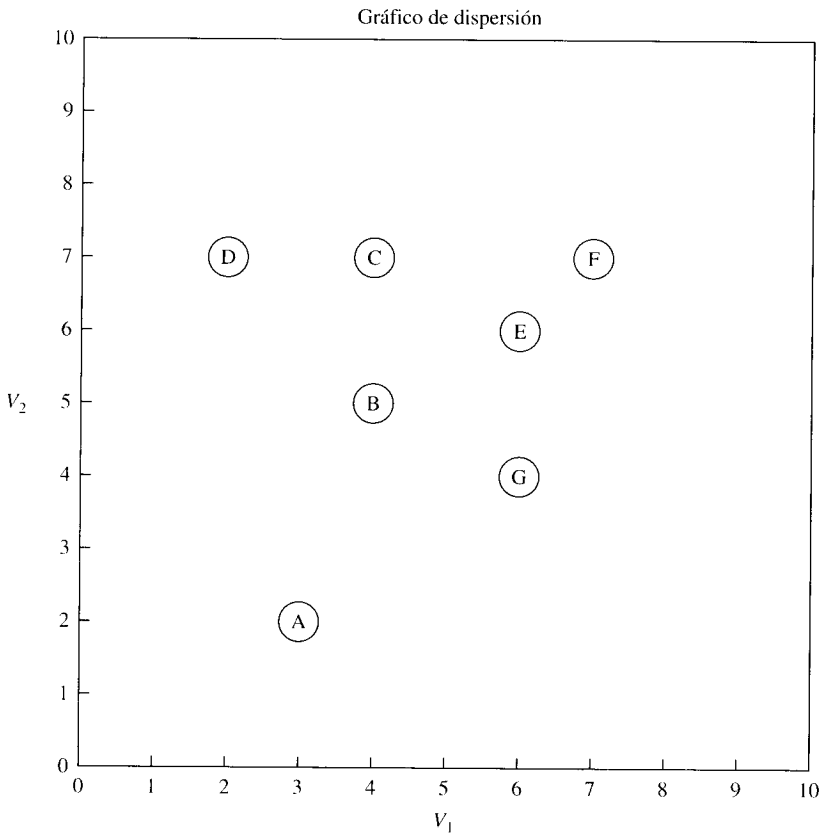


FIGURA 9.1. Valores de los datos y gráfico de dispersión para las siete observaciones basados en las dos variables de aglomeración (V_1, V_2).

mia, se refleja en el menor número de conglomerados posible. Pero a medida que el número de conglomerados disminuye, la homogeneidad dentro de los conglomerados necesariamente disminuye. Por tanto, se debe buscar un equilibrio entre la definición de las estructuras más básicas (pocos conglomerados) que todavía mantienen el necesario nivel de similitud dentro de los conglomerados. Una vez que tengamos procedimientos para cada asunto, podemos realizar el análisis *cluster*.

Medición de la similitud

Ilustraremos un análisis *cluster* para las siete observaciones (encuestados A-G) utilizando procedimientos sencillos para cada uno de los asuntos. En primer lugar, la similitud será medida de acuerdo con la **distancia Euclídea** (en línea recta) entre cada par de observaciones. La Tabla 9.1 contiene medidas de proximidad entre cada uno de los siete encuestados. Al utilizar la distancia como medida de proximidad, debemos recordar que las distancias más pequeñas indican mayor similitud, de tal forma que las observaciones E y F son las más parecidas (1,414), y A y F son las más diferentes (6,403).

TABLA 9.1. Matriz de proximidad de distancias euclídeas entre observaciones

	Observación						
Observación	*A*	*B*	*C*	*D*	*E*	*F*	*G*
A	—						
B	3,162	—					
C	5,099	2,000	—				
D	5,099	2,828	2,000	—			
E	5,000	2,236	2,236	4,123	—		
F	6,403	3,606	3,000	5,000	1,414	—	
G	3,606	2,236	3,606	5,000	2,000	3,162	—

Formación de conglomerados

Una vez tenemos la medida de similitud, debemos desarrollar el siguiente procedimiento para la formación de conglomerados. Como se mostrará posteriormente, se han propuesto muchos métodos. Pero para nuestros propósitos, utilizamos esta regla simple: identificar las dos observaciones más parecidas (cercanas) que no están en el mismo conglomerado y combinar éstas. Aplicamos esta regla repetidas veces, comenzando con cada observación en su propio «conglomerado» y combinando dos conglomerados a un tiempo hasta que todas las observaciones estén en un único conglomerado. A esto se le denomina un **procedimiento jerárquico** dado que se opera paso a paso para formar un rango completo de soluciones *cluster*. Es también un **método aglomerativo** dado que los conglomerados se forman por la combinación de los conglomerados existentes.

La Tabla 9.2 detalla los pasos del procedimiento jerárquico, en primer lugar representando el estado inicial con las siete observaciones en conglomerados simples. A continuación se unen los conglomerados en el proceso aglomerativo hasta que sólo quede un conglomerado. El paso 1 identifica las dos observaciones más cercanas (E y F) y las combina en un conglomerado, yendo de siete a seis conglomerados. A continuación, el paso 2 busca los pares de observaciones más cercanos. En este caso, tres pares tiene la misma distancia de 2,000 (E-G, C-D, y B-C). Comencemos con E-G. G es un miembro único de un conglomerado, pero E se combinó en el primer paso con F. Así, el conglomerado formado a este nivel tiene tres miembros: G, E y F. El paso 3 combina los conglomerados de miembro único de C y D y el paso 4 combina B con el conglomerado de dos miembros C-D que se formó en el paso 3. Hasta este momento, tenemos 3 conglomerados: conglomerado 1 (A), conglomerado 2 (B, C y D), y conglomerado 3 (E, F y G).

TABLA 9.2. Proceso de *cluster* aglomerativo jerárquico

	Proceso de aglomeración		*Solución Cluster*		
Paso	*Distancia mínima entre observaciones conjunta (distancias medias no aglomeradas[a])*	*Par de Observaciones*	*Pertenencia al conglomerado*	*Número de conglomerados*	*Medida de similitud número de conglomerados (dentro del conglomerado)*
	Solución inicial		*(A) (B) (C) (D) (E) (F) (G)*	7	0
1	1,414	E-F	(A) (B) (C) (D) (E-F) (G)	6	1,414
2	2,000	E-G	(A) (B) (C) (D) (E-F–G)	5	2,192
3	2,000	C-D	(A) (B) (C–D) (E-F–G)	4	2,144
4	2,000	B-C	(A) (B–C-D) (E-F-G)	3	2,234
5	2,236	B-E	(A) (B-C-D–E-F-G)	2	2,896
6	3,162	A-B	(A–B-C-D-E-F-G)	1	3,420

[a]Distancia euclídea entre observaciones.

La siguiente distancia más pequeña es 2,236 para tres pares de observaciones (E-B, B-G y C-E). Utilizamos sólo una de estas tres distancias, sin embargo, en la medida en que cada par de observaciones contiene un miembro de cada uno de los dos conglomerados existentes (B, C y D frente a E, F y G). Por tanto, el paso 5 combina los dos conglomerados de tres miembros en un único conglomerado de seis miembros. El paso final (6) es combinar la observación A con el conglomerado restante (seis observaciones) en un único conglomerado a una distancia de 3,162. Notará que existen tres distancias iguales o menores a 3,162, pero que no se utilizan porque están entre los miembros del mismo conglomerado.

El proceso jerárquico de aglomeración puede representarse gráficamente de varias formas. La Figura 9.2 ilustra dos de tales métodos. En primer lugar, dado que el proceso es jerárquico, el proceso de aglomeración puede mostrarse como series de agrupaciones anidadas (véase Figura 9.2a). Este proceso, sin embargo, puede representar la proximidad de las observaciones para sólo dos o

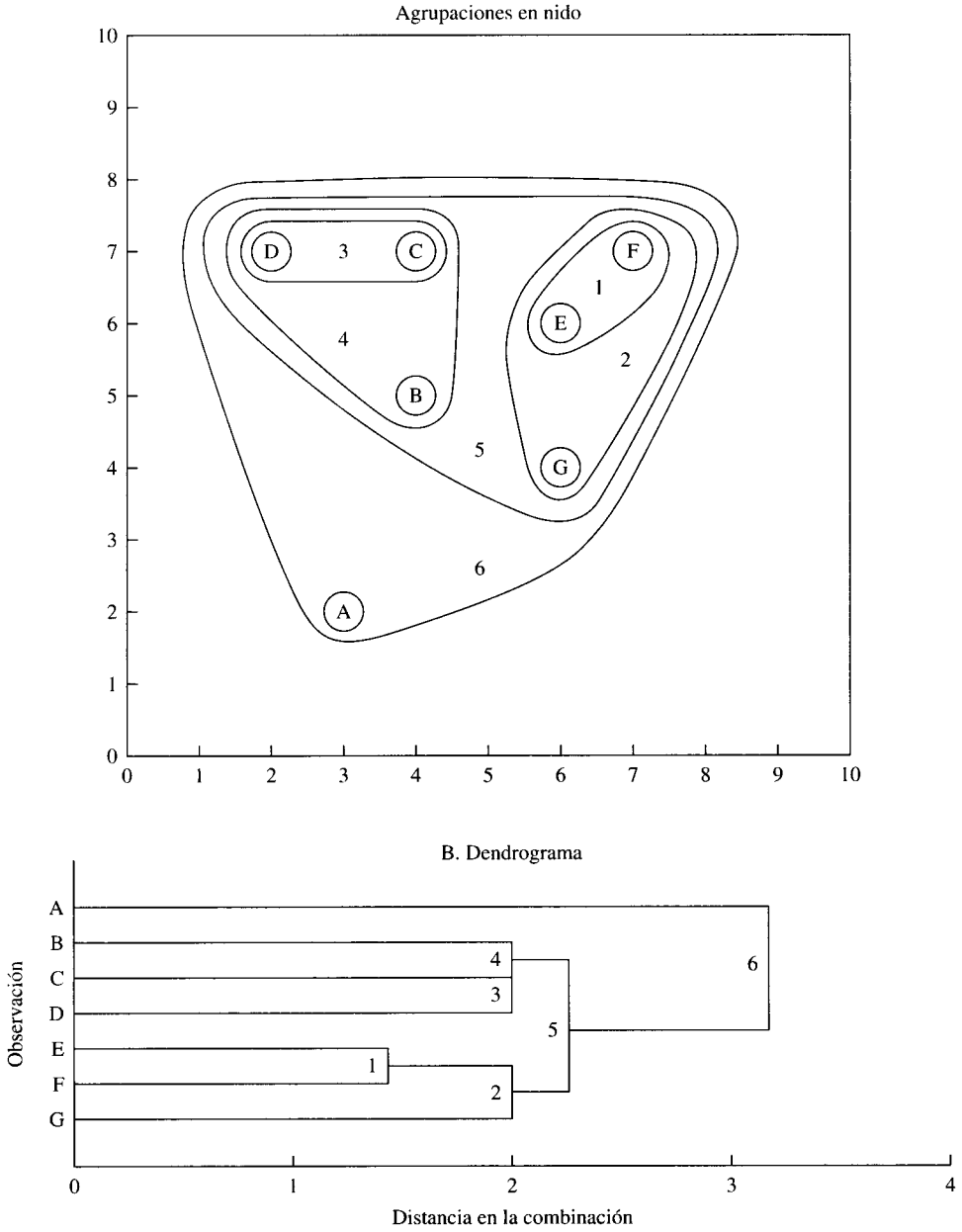


FIGURA 9.2. Representaciones gráficas del proceso de aglomerativo jerárquico

tres variables de aglomeración del gráfico tridimensional o de dispersión. Una aproximación más habitual es el dendrograma, que representa el proceso de aglomeración en un gráfico con forma de árbol. El eje horizontal representa el coeficiente de aglomeración, en este caso la distancia utilizada en la unión de aglomerados. Esta aproximación es particularmente útil en la identificación de atípicos, como la observación A. También representa el tamaño relativo de los conglomerados que varían, aunque se hace difícil de manejar cuando aumenta el número de observaciones.

Determinación del número de conglomerados en la solución final

Un método jerárquico produce un número de soluciones *cluster* —en este caso van de una solución de un conglomerado a una solución de seis conglomerados. Pero ¿cuál deberíamos elegir? Sabemos que a medida que nos alejamos de los conglomerados de un único miembro, la homogeneidad disminuye. Así que, ¿por qué no quedarnos con los siete conglomerados, que son los más homogéneos posible? El problema es que no hemos definido ninguna estructura con siete conglomerados. Así que el investigador debe ver cada solución *cluster* a partir de la descripción de su estructura compensada con la homogeneidad de los conglomerados. En este ejemplo, utilizamos una medida muy simple de homogeneidad: las distancias medias de todas las observaciones dentro de los conglomerados. En la solución inicial con siete conglomerados, nuestra medida de similitud conjunto es 0 —ninguna observación está emparejada con otra. Para la solución de seis conglomerados, la similitud conjunta es la distancia entre las dos observaciones (1,414) unidas en el paso 1. El paso 2 forma un conglomerado de tres miembros (E, F y G), de tal forma que la medida de similitud total es la media de las distancias entre E y F (1,414), E y G (2,000), y F y G (3,162), para una media de 2,192. En el paso 3, se forma un nuevo conglomerado de dos miembros con una distancia de 2,000, que provoca que la media conjunta caiga ligeramente hasta 2,144. Podemos proceder a formar nuevos conglomerados de esta forma hasta formar una solución de conglomerado único (paso 6), en el que la media de todas las distancias de la matriz de distancias es 3,420.

Ahora bien, ¿cómo utilizamos esta medida conjunta de similitud para seleccionar una solución *cluster*? Recordemos que estamos intentando conseguir la estructura más simple posible que represente agrupaciones homogéneas. Si controlamos la medida de similitud conjunta a medida que disminuye el número de conglomerados, grandes aumentos en la medida conjunta indican que dos conglomerados no eran tan similares. En nuestro ejemplo, la medida conjunta aumenta cuando en primer lugar juntamos dos observaciones (paso 1) y a continuación lo hacemos de nuevo cuando construimos nuestro primer conglomerado de tres miembros (paso 2). Pero en los dos pasos siguientes (3 y 4), la medida conjunta no cambia substancialmente. Esto indica que estamos formando otros conglomerados prácticamente con la misma homogeneidad de los conglomerados existentes. Pero cuando alcanzamos el paso 5, que combina los dos conglomerados de tres miembros, observamos un gran aumento. Esto indica que al unir estos dos conglomerados obtenemos un único conglomerado marcadamente menos homogéneo. Consideramos la solución *cluster* del paso 4 mucho mejor que la del paso 5. Podemos también ver que en el paso 6 la medida conjunta de nuevo aumenta ligeramente indicando que, incluso aunque la última observación permanezca separada hasta el último paso, cuando se une cambia la homogeneidad del conglomerado. Sin embargo, dado el perfil bastante aislado de la observación A comparada con el resto, puede ser mejor designar como miembro del **grupo de entropía**, aquellas observaciones que son atípicos e independientes de los conglomerados existentes. Por tanto, cuando se revisa el rango de las soluciones *cluster*, la solución de tres conglomerados del paso 4 parece la más apropiada para una solución *cluster* definitiva, con dos conglomerados de igual tamaño y una única observación atípica.

Como ya habrá quedado claro, en la selección de la solución *cluster* definitiva se deja al juicio del observador y es considerado por muchos como un proceso muy subjetivo. Incluso aunque se han desarrollado métodos más sofisticados para ayudar en la evaluación de las soluciones *cluster*, sigue recayendo en el investigador la decisión final del número de conglomerados aceptados

en la solución final. El análisis *cluster* es más simple en este caso bivariante porque los datos están en dos dimensiones. En la mayoría de los estudios de márketing, sin embargo, se miden más de dos variables con cada objeto, y la situación es mucho más compleja con muchas más observaciones. En lo que resta de capítulo, discutiremos cómo el investigador puede emplear procedimientos más sofisticados para tratar con el aumento de la complejidad de las aplicaciones al «mundo real».

Proceso de decisión con el análisis *cluster*

El análisis *cluster*, como el resto de las técnicas multivariantes discutidas anteriormente, puede verse como una aproximación a la construcción de modelos en seis pasos incluido en el Capítulo 1 (véase Figura 9.3 para los pasos 1-3 y Figura 9.6 para los pasos 4-6). Comenzando con los objetivos de investigación que pueden ser tanto confirmatorios como exploratorios, el diseño de un análisis *cluster* interviene en la partición del conjunto de datos para formar conglomerados, la interpretación de los conglomerados y la validación de los resultados. El proceso de participación determina cómo se pueden desarrollar los conglomerados. El proceso de interpretación implica entender las características de cada conglomerado y desarrollar un nombre o etiqueta que defina apropiadamente su naturaleza. El proceso final comprende la evaluación de la validación de la solución *cluster* (es decir, determinación de su estabilidad y generalidad), junto con la descripción de las características de cada conglomerado para explicar cómo puede diferir en dimensiones relevantes como las demográficas. Las siguientes secciones detallan todos estos asuntos a lo largo de un proceso de construcción de modelos de seis pasos.

Primer paso: Objetivos del análisis *cluster*

El objetivo fundamental del análisis *cluster* es la obtención de un conjunto de objetos en dos o más grupos basándose en su similitud para un conjunto de características especificadas (valor teórico del análisis *cluster*). Al formar grupos homogéneos, el investigador puede conseguir los siguientes objetivos:

1. *Descripción de una taxonomía*. El uso más tradicional del análisis *cluster* ha sido para propósitos exploratorios y la formulación de una **taxonomía** —una clasificación de objetos realizada empíricamente. Como se ha descrito previamente, el análisis *cluster* se ha utilizado para un amplio rango de aplicaciones debido a su capacidad para la partición. Pero el análisis *cluster* puede generar también hipótesis relacionadas con la estructura de los objetos. Sin embargo, aunque visto principalmente como una técnica de exploración, el análisis *cluster* puede utilizarse a efectos confirmatorios. Si una estructura propuesta puede definirse para un conjunto de objetos, se puede aplicar el análisis *cluster*, y puede compararse una **tipología** propuesta (clasificación basada en la teoría) a la derivada del análisis *cluster*.
2. *Simplificación de los datos*. En el curso de la obteción de una taxonomía, el análisis *cluster* también obtiene una perspectiva simplificada de las observaciones. Con una estructura definida, las observaciones pueden agruparse para análisis ulteriores. Mientras el análisis factorial intenta proporcionar «dimensiones» o estructuras de variables (véase Capítulo 3), el análisis *cluster* desarrolla la misma tarea para las observaciones. Por tanto, en lugar de ver todas las observaciones como únicas, pueden ser consideradas como miembros de un conglomerado y perfiladas por sus características generales.
3. *Identificación de relación*. Con los conglomerados definidos y la estructura subyacente de los datos representada en dichos conglomerados, el investigador tiene un medio de reve-

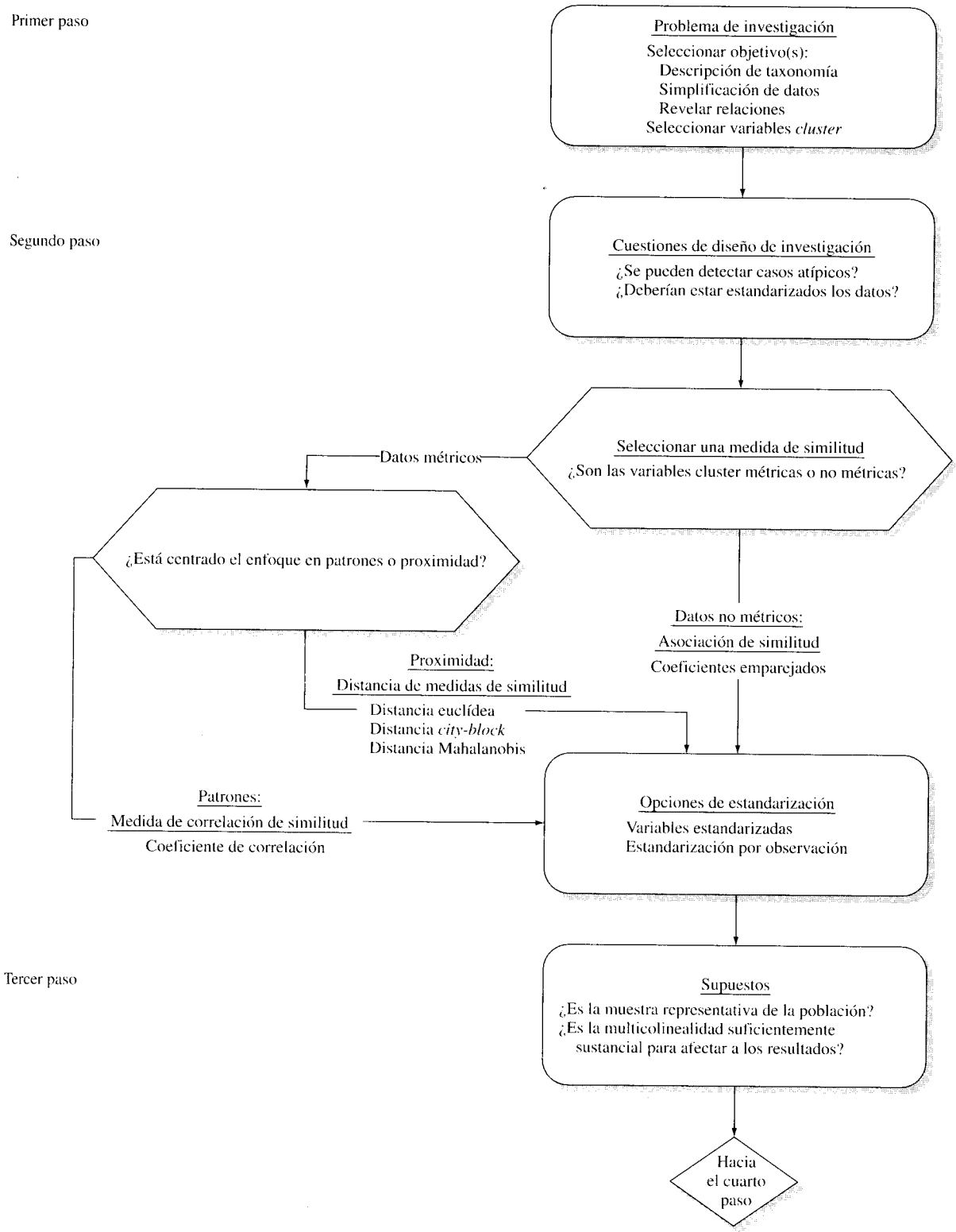


FIGURA 9.3. Diagrama de decisión de los pasos 1-3 del análisis *cluster*.

lar relaciones entre las observaciones que quizá no fuese posible con las observaciones individuales. Mientras se utilizan análisis tales como el discriminante para identificar relaciones empíricamente, o los grupos están sujetos a métodos más cualitativos, la estructura simplificada del análisis *cluster* muchas veces representa relaciones o similitudes y diferencias no reveladas previamente.

Selección de variables del análisis *cluster*

En cualquier aplicación, los objetivos del análisis *cluster* no pueden separarse de la selección de variables utilizadas para caracterizar los objetos a agrupar. Tanto si el objetivo es exploratorio como confirmatorio, el investigador ha restringido efectivamente los resultados posibles por las variables elegidas para el uso. Los conglomerados derivados reflejan la estructura inherente de los datos sólo como definida por las variables.

La selección de las variables a incluir en el valor teórico del análisis *cluster* debe hacerse con relación a consideraciones teóricas, conceptuales y prácticas. Cualquier aplicación del análisis *cluster* debe descansar en cierta lógica en función de la cual se seleccionan las variables. Tanto si dicha lógica se basa en una teoría explícita, investigación pasada o suposición, el investigador debe darse cuenta de la importancia de incluir sólo aquellas variables que (1) caracterizan los objetos que se están agrupando, y (2) se refieren específicamente los objetivos del análisis *cluster*. La técnica del análisis *cluster* no tiene un medio para diferenciar las variables relevantes de las irrelevantes. Sólo obtiene los grupos de objetos más consistentes, aunque diferenciados, para *todas* las variables. La inclusión de una variable irrelevante aumenta la posibilidad de que se creen atípicos sobre estas variables, que pueden tener un efecto importante sobre los resultados. Por tanto, uno nunca debería incluir variables indiscriminadamente sino en su lugar elegir las variables utilizando el objetivo de investigación como criterio de selección.

A efectos prácticos, el análisis cluster puede verse drásticamente afectado por la inclusión de una o dos variables inapropiadas o escasamente diferenciadas [8]. Se anima al investigador a examinar los resultados y eliminar las variables que no son distintivas (es decir, que no difieren significativamente) de todos los conglomerados deducidos. Este procedimiento permite a las técnicas *cluster* maximizar los conglomerados definidos basándose sólo en aquellas variables que exhiban diferencias para todos los objetos.

Segundo paso: Diseño de investigación mediante análisis *cluster*

Con los objetivos definidos y las variables seleccionadas, el investigador debe tratar tres cuestiones antes de empezar el proceso de partición: (1) ¿pueden detectarse los atípicos y, si es posible, deberían ser destruidos? (2) ¿cómo debería medirse la similitud de los objetos? y (3) ¿deberían estandarizarse los datos? Se pueden utilizar muchos enfoques para contestar a estas preguntas. Sin embargo, ninguno de ellos ha sido evaluado suficientemente como para ofrecer una respuesta definitiva a cualquiera de estas cuestiones, y, desafortunadamente, muchas de las aproximaciones ofrecen diferentes resultados para el mismo conjunto de datos. Por tanto, el análisis *cluster* junto con el análisis factorial, es más un arte que una ciencia. Por esta razón, nuestra discusión revisa estos supuestos de forma muy general ofreciendo ejemplos de los enfoques habitualmente más utilizados y una evaluación de las limitaciones prácticas siempre que sea posible.

La importancia de estos supuestos y las decisiones hechas en los últimos pasos se hacen aparentes cuando nos damos cuenta de que aunque el análisis *cluster* está buscando una estructura de los datos, debe imponer en realidad una estructura a partir de una metodología seleccionada. El análisis *cluster* no puede evaluar todas las posibles particiones porque, incluso para un problema relativamente pequeño de partición de 25 objetos en 5 conglomerados no solapados, existen $2{,}4 \times 10^{15}$

particiones posibles [2]. En su lugar, basándose en las decisiones del investigador, la técnica identifica una de las posibles situaciones como «correcta». Desde este punto de vista, los supuestos del diseño de investigación y la elección de metodologías hecha por el investigador tienen quizá un impacto superior al del resto de las técnicas multivariantes.

Detección de atípicos

En su búsqueda de una estructura, el análisis *cluster* es muy sensible a la inclusión de variables irrelevantes. Pero el análisis *cluster* es también sensible a los atípicos (objetos que son muy diferentes del resto). Los atípicos pueden representar tanto (1) observaciones verdaderamente «aberrantes» que no son representativas de la población en general, o (2) una muestra reducida del grupo (grupos) de la población que provoca una mala representación del grupo (grupos) de la muestra. En ambos casos, los atípicos distorsionan la verdadera estructura y hacen que los conglomerados deducidos no sean representativos de la verdadera estructura de la población. Por esta razón, siempre es necesaria una representación preliminar de los atípicos. Probablemente la forma más sencilla de llevar a cabo esta representación es preparar un diagrama de perfil gráfico, tal como se muestra en la Figura 9.4. El diagrama de perfil tiene las variables en el eje horizontal y los valores de las variables a lo largo del eje vertical. Cada punto del gráfico representa el valor de la correspondiente variables, y los puntos están conectados para facilitar la interpretación visual. Se representan en el gráfico los perfiles para todos los objetos, una línea para cada objeto. Los atípicos son aquellos objetos con perfiles muy diferentes, la mayoría caracterizados por valores extremos sobre una o más variables.

Obviamente, tal procedimiento se hace incómodo con un gran número de objetos (observaciones) o variables. Para las observaciones que se muestran en la Figura 9.4, no hay un atípico obvio que tenga valores extremadamente bajos o altos. Pero al igual que en la detección de atípicos multivariantes utilizados en otras técnicas multivariantes, los atípicos pueden definirse también a partir de perfiles únicos que los distinguen del resto de las observaciones. Para estos casos, se pueden aplicar los procedimientos de identificación de atípicos discutidos en el Capítulo 2. También, pueden surgir en el cálculo de similitud. Cualquiera que sea el medio utilizado, las observaciones identificadas como atípicos puede evaluarse a efectos de su representatividad respecto de la población y eliminarlos del análisis si se consideran no representativos. Pero, como en otros casos de detección de atípicos, el investigador debería tener precaución en la eliminación de observaciones de la muestra porque tal eliminación puede distorsionar la estructura efectiva de los datos.

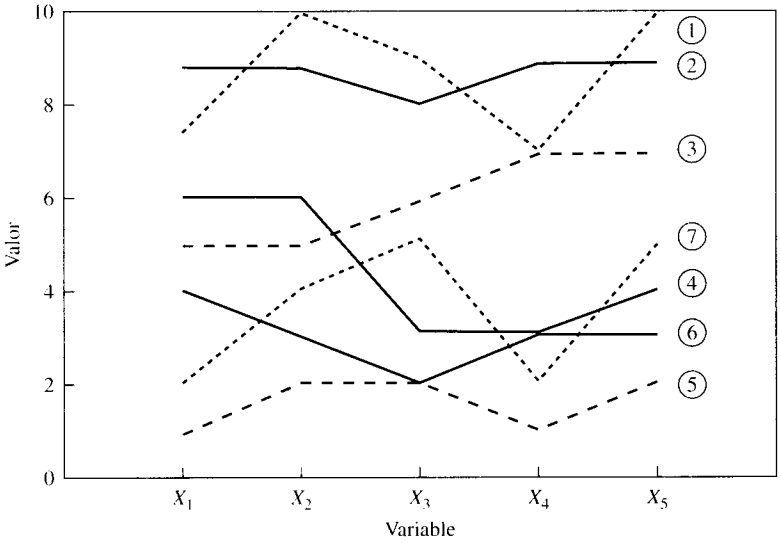


FIGURA 9.4. Diagrama de perfil.

Medidas de similitud

El concepto de similitud es fundamental para el análisis *cluster*. La **similitud entre objetos** es una medida de correspondencia, o parecido, entre objetos que van a ser agrupados. En nuestra discusión del análisis factorial, creamos una matriz de correlación entre las variables que fueron utilizadas para agrupar las variables en factores. Un proceso comparable se produce en el análisis *cluster*. Aquí, las características que definen la similitud se especifican en primer lugar. A continuación, se combinan las características en una medida de similitud calculada para todos los pares de objetos, al igual que utilizábamos las correlaciones entre variables en el análisis factorial. De esta forma, cualquier objeto puede ser comparado para cualquier objeto a través de una medida de similitud. El procedimiento del análisis *cluster* procede a continuación a agrupar objetos similares juntos en los conglomerados.

La similitud entre objetos puede medirse de varias formas, pero tres métodos dominan las aplicaciones del análisis *cluster*: medidas de correlación, medidas de distancia y medidas de asociación. Cada uno de los métodos representa una perspectiva particular de similitud, dependiendo tanto de sus objetivos como del tipo de datos. Tanto las medidas de distancia como la correlación exigen datos métricos, mientras que las medidas de asociación son para datos no métricos.

Medidas de correlación

La medida de similitud entre objetos que probablemente nos viene a la mente en primer lugar es el coeficiente de correlación entre un par de objetos medido sobre varias variables. En efecto, en lugar de hacer la correlación entre dos conjuntos de variables, invertimos la matriz de las X variables de los objetos de tal forma que las columnas representan los objetos y las filas representan las variables. Por tanto, el coeficiente de correlación entre las dos columnas de números es la correlación (o similitud) entre los perfiles de los dos objetos. Elevadas correlaciones indican similitud y bajas correlaciones indican falta de ella. Este procedimiento se sigue en la aplicación del análisis de factor de tipo Q (véase Capítulo 3).

Las medidas de correlación representan la similitud mediante la correspondencia de patrones entre las características (X variables). Esto se ilustra por el ejemplo de siete observaciones que se muestra en la Figura 9.4. Una medida de correlación de similitudes no observa las magnitudes sino las patrones de los valores. En la Tabla 9.3, que contiene las correlaciones entre estas siete observaciones, podemos ver dos grupos distintos. En primer lugar, los casos 1, 5 y 7 tienen patrones similares y corresponden a intercorrelaciones positivas elevadas. De la misma forma, los casos 2, 4 y 6 también tienen correlaciones positivas elevadas entre ellos pero con correlaciones bajas o nulas con las otras observaciones. El caso 3 tiene correlaciones bajas o negativas con el resto de los casos, por lo que quizá forme un grupo en sí mismo. Por tanto, las correlaciones representan patrones para todas las variables más que las magnitudes. Las medidas de correlación, sin embargo, se utilizan rara vez porque el interés de la mayoría de las aplicaciones del análisis *cluster* está en las magnitudes de los objetos, no en los patrones de los valores.

Medidas de distancia

Incluso aunque las medidas de correlación tienen un atractivo intuitivo y se utilizan en otras muchas técnicas multivariantes, no son las medidas de similitud más utilizadas en el análisis *cluster*. Las medidas de similitud de distancia, que representan la similitud como la proximidad de las observaciones respecto a las otras para las variables del valor teórico del análisis *cluster*, son las medidas de similitud más utilizadas. Las medidas de distancia son en realidad medidas de diferencia, donde los valores elevados indican una menor similitud. La distancia se convierte en medida de similitud utilizando una relación inversa. En nuestro hipotético ejemplo se muestra una ilustración simple de esta relación, en la que los conglomerados de observaciones se definieron de acuerdo a la proximidad de unas observaciones respecto de otras, cuando se representa gráficamente las puntuaciones de las observaciones sobre dos variables (véase Figura 9.2).

TABLA 9.3. Cálculo de las medidas de similitud de distancia y correlación

DATOS ORIGINALES

	Variables				
Caso	X_1	X_2	X_3	X_4	X_5
1	7	10	9	7	10
2	9	9	8	9	9
3	5	5	6	7	7
4	6	6	3	3	4
5	1	2	2	1	2
6	4	3	2	3	3
7	2	4	5	2	5

MEDIDA DE SIMILITUD: CORRELACIÓN

	Caso						
Caso	1	2	3	4	5	6	7
1	1,00						
2	−0,147	1,00					
3	0,000	0,000	1,00				
4	0,087	0,516	−0,824	1,00			
5	0,963	−0,408	0,000	−0,060	1,00		
6	−0,466	0,791	−0,354	0,699	−0,645	1,00	
7	0,891	−0,516	0,165	−0,239	0,963	−0,699	1,00

MEDIDA DE SIMILITUD: DISTANCIA EUCLÍDEA

	Caso						
Caso	1	2	3	4	5	6	7
1	nc						
2	3,32	nc					
3	6,86	6,63	nc				
4	10,24	10,20	6,00	nc			
5	15,78	16,19	10,10	7,07	nc		
6	13,11	13,00	7,28	3,87	3,87	nc	
7	11,27	12,16	6,32	5,10	4,90	4,36	nc

nc = distancia no calculada.

Comparación con las medidas de correlación Las diferencias entre las medidas de distancia y de correlación pueden verse volviendo a la Figura 9.4. Las medidas de distancia se centran en la magnitud de los valores y representan casos similares que están juntos, pero que tiene pautas muy distintas para todas las variables. La Tabla 9.3 también contiene las medidas de distancia de similitud para los siete casos, y observamos casos de aglomeraciones muy diferentes surgiendo de aquellos encontrados cuando se utilizan las medidas de correlación. Con las distancias más reducidas representando similitud, observamos que los casos 1 y 2 forman un grupo, y los casos 4, 5, 6 y 7 otro grupo. Estos grupos representan aquellos con valores más altos frente a los más bajos. Un tercer grupo, que consiste sólo en el caso 3, difiere de los otros porque tiene valores que son tanto bajos como altos. Aunque los dos conglomerados que utilizan las medidas de distancia tienen diferentes miembros que aquellos que utilizan las correlaciones, el caso 3 es único en cada medida de similitud. La elección de una medida de correlación en lugar de la medida más tradicional de distancia requiere una interpretación muy diferente de los resultados por el investigador. Los conglomerados basados en medidas de correlación pueden no tener valores similares en lugar de tener patrones similares. Los conglomerados basados en la distancia tienen valores más parecidos para el conjunto de variables, pero los patrones pueden ser bastante diferentes.

Tipos de medidas de distancia Existen varias medidas de distancia. La más utilizada es la **distancia Euclídea**. Un ejemplo de cómo se obtiene la distancia Euclídea se muestra geométricamente en la Figura 9.5. Suponga que dos puntos de dos dimensiones tienen coordenadas (X_1, Y_1) y (X_2, Y_2), respectivamente. La distancia Euclídea entre los puntos es la longitud de la hipotenusa de un triángulo rectángulo, calculada por la fórmula que está bajo la figura. Este concepto es fácilmente generalizable para más de dos variables. La distancia Euclídea se utiliza para calcular medidas específicas tales como la simple distancia Euclídea (calculada como se ha descrito antes) y la distancia Euclídea cuadrada, o absoluta, que es la suma de las diferencias al cuadrado sin tomar la raíz cuadrada. La distancia Euclídea al cuadrado tiene la ventaja de no tener que tomar la raíz cuadrada lo que acelera notablemente los cálculos, y es la medida de distancia recomendada para los métodos de análisis *cluster* del centroide y Ward.

Hay varias opciones que no se basan en la distancia Euclídea. Una de las medidas alternativas más utilizadas consiste en reemplazar la diferencia de los cuadrados por la suma de las diferencias absolutas de las variables. Este procedimiento se denomina función de la distancia absoluta. El **enfoque de la distancia absoluta** puede resultar apropiado bajo ciertas circunstancias, pero puede provocar varios problemas. Uno es el supuesto de que las variables no están correlacionadas con el resto; si lo están, los conglomerados no son válidos [15]. En la mayoría de los programas del análisis *cluster* se encuentran otras medidas que emplean variaciones de las diferencias absolutas o las potencias aplicadas a las diferencias (que no sean sólo la diferencia de los cuadrados).

Impacto de los valores de los datos no estandarizados Un problema al que se enfrentan todas las medidas de distancia es que el uso de datos no estandarizados implica inconsistencias entre las soluciones *cluster* cuando cambia la escala de las variables. Por ejemplo, suponga que tres objetos, A, B y C se miden sobre dos variables, probabilidad de compra de la marca X (en porcentajes) y cantidad de tiempo gastado viendo anuncios de la marca X (en minutos o segundos). Los valores de cada observación se muestran en la Tabla 9.4.

Con esta información, se pueden calcular las medidas de distancia. En nuestro ejemplo, calculamos tres medidas de distancia para cada par de objetos: distancia Euclídea simple, la distancia Euclídea absoluta o al cuadrado y la distancia absoluta. En primer lugar, calculamos los valores de las distancias de acuerdo con la probabilidad de compra y tiempo de visión en minutos. En la Tabla 9.4 se muestran estas distancias, donde los valores más reducidos indican mayor proximidad y similitud, así como su puesto de orden. Como podemos ver, los objetos más parecidos (con la menor distancia) son B y C, seguidos de A y C, con A y B siendo los menos parecidos (menos próximos). Este orden se mantiene para las tres medidas de distancia, pero la similitud relativa o dispersión entre los objetos está más pronunciada en la medida de distancia Euclídea al cuadrado.

El orden de las similitudes puede cambiar profundamente con sólo un cambio en la escala de una de las variables. Si medimos el tiempo de visión en segundos en lugar de en minutos, enton-

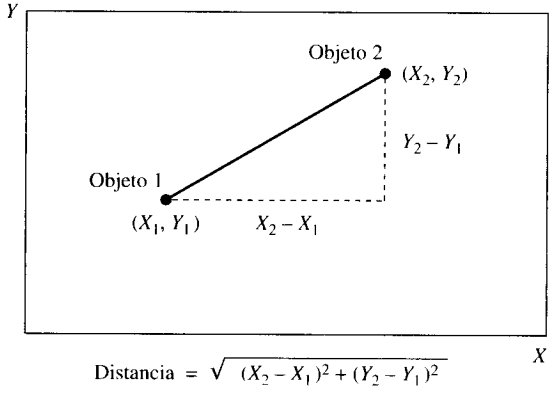


FIGURA 9.5. Un ejemplo de distancia euclídea entre dos objetos medidos sobre dos variables, X e Y.

TABLE 9.4. Variaciones en las medidas de distancia basads en distintas escalas de medida

DATOS ORIGINALES

Objeto	*Probabilidad de compra*	*Tiempo de visión del anuncio*	
		Minutos	*Segundos*
A	60	3,0	180
B	65	3,5	210
C	63	4,0	240

MEDIDAS DE DISTANCIA BASADAS EN LA PROBABILIDAD DE COMPRA Y TIEMPO EN MINUTOS DE VISIÓN DE ANUNCIOS

	Distancia Euclídea simple		*Distancia Euclídea absoluta o cuadrado*		*Distancia City-block*	
Par de objetos	*Valor*	*Rango*	*Valor*	*Rango*	*Valor*	*Rango*
A-B	5,025	3	25,25	3	5,5	3
A-C	3,162	2	10,00	2	4,0	2
B-C	2,062	1	4,25	1	2,5	1

MEDIDAS DE DISTANCIA BASADAS EN LA PROBABILIDAD DE COMPRA Y TIEMPO EN SEGUNDOS DE VISIÓN DE ANUNCIOS

	Distancia Euclídea simple		*Distancia Euclídea absoluta o cuadrado*		*Distancia City-block*	
Par de objetos	*Valor*	*Rango*	*Valor*	*Rango*	*Valor*	*Rango*
A-B	30,41	2	925	2	35	3
A-C	60,07	3	3,609	3	63	2
B-C	30,06	1	904	1	32	1

MEDIDAS DE DISTANCIA BASADAS EN VALORES ESTANDARIZADOS DE PROBABILIDAD DE COMPRA Y TIEMPO EN MINUTOS O SEGUNDOS DE VISIÓN DE ANUNCIOS

	Valores estandarizados		*Distancia Euclídea simple*		*Distancia Euclídea absoluta o cuadrado*		*Distancia City-block*	
Par de objetos	*Probabilidad de compra*	*Minutos/segundos de ver anuncio*	*Valor*	*Rango*	*Valor*	*Rango*	*Valor*	*Rango*
A-B	−1,06	−1,0	2,22	2	4,95	2	2,99	2
A-C	0,93	0,0	2,33	3	5,42	3	3,19	3
B-C	0,13	1,0	1,28	1	1,63	1	1,79	1

ces los órdenes cambian (véase Tabla 9.4). Los objetos B y C siguen siendo los más parecidos, pero ahora el par A-B es más parecido y casi idéntico a la similitud de B-C. Pero cuando tenenos minutos de tiempo de visión, el par A-B es el menos parecido por un margen substancial. Lo que ha ocurrido es que la escala de tiempo de visión variable ha dominado los cálculos, haciendo la probabilidad de compra menos significativa en los cálculos. Y al contrario, cuando medimos el tiempo en minutos, es entonces la probabilidad de compra la que domina en los cálculos. El investigador notará el tremendo impacto que la escala de las variables puede tener sobre la solución final. Debería emplearse la estandarización de las variables de aglomeración, siempre que sea conceptualmente posible, para evitar casos como los encontrados en nuestro ejemplo.

Una medida de distancia Euclídea habitualmente utilizada que incorpora directamente un procedimiento de estandarización es la **distancia de Mahalanobis (D^2)**. El enfoque de Mahalanobis no sólo realiza el proceso de estandarización de los datos a escala en términos de las desviaciones estándar sino que también evalúa la varianza-covarianza unidas dentro del grupo, que ajusta las intercorrelaciones entre las variables. Conjuntos de variables altamente intercorrelacionados del análisis *cluster* pueden ponderar implícitamente un conjunto de variables en los procedimientos de aglomeración. En resumen, el procedimiento de distancia generalizada de Mahalanobis calcula una

medida de distancia entre objetos comparable al R^2 del análisis de regresión. Aunque en muchas situaciones puede ser apropiado el uso de la distancia de Mahalanobis, no todos los programas la incluyen como medida de similitud. En tales casos, el investigador normalmente selecciona la distancia Euclídea al cuadrado.

Al intentar seleccionar una medida de distancia particular, el investigador debería recordar las siguientes advertencias. Diferentes medidas de distancia o un cambio en la escala de las variables pueden llevar a diferentes soluciones *cluster*. Por tanto, es aconsejable utilizar varias medidas y comparar los resultados con pautas teóricas o conocidas. También, cuando las variables están intercorrelacionadas (sea positiva o negativa), la medida de distancia de Mahalanobis es probable que sea la más apropiada dado que se ajusta para las correlaciones y ponderaciones de todas las variables igualmente. Desde luego, si el investigador desea ponderar las variables desigualmente, existen otros procedimientos [10, 11].

Medidas de asociación

Las medidas de asociación de similitud se utilizan para comparar objetos cuyas características se miden sólo en términos no métricos (medida nominal y ordinal). Como ejemplo, veamos el caso en que los encuestados responden sí o no a cierto número de preguntas. Una medida de asociación podría evaluar el grado de acuerdo o de acercamiento entre cada par de encuestados. La forma más simple de medida de asociación sería el porcentaje de veces que existió acuerdo (ambos dicen sí o ambos dicen no) para el mismo conjunto de cuestiones. Se han desarrollado extensiones de este simple coeficiente de ajuste para acomodar variables nominales de varias categorías o incluso medidas ordinales. Muchos programas informáticos, sin embargo, dan un apoyo limitado a las medidas de asociación, y el investigador está en muchas ocasiones forzado a calcular en primer lugar las medidas de similitud y a continuación introducir la matriz de similitud en los programas de análisis *cluster*. Hay diversas fuentes en las que se pueden encontrar revisiones de varios tipos de medidas de asociación [4, 16].

Tipificación de los datos

Con la medida de similitud seleccionada, el investigador debe tratar sólo con una cuestión más: ¿deberían tipificarse los datos antes de calcular las similitudes? Al contestar esta cuestión, el investigador debe considerar varios asuntos. En primer lugar, la mayoría de las distancias medidas son bastantes sensibles a las diferentes escalas o magnitudes de las variables. Vimos este impacto previamente cuando cambiamos de minutos a segundos en una de nuestras variables. En general, las variables con una mayor dispersión (es decir, grandes desviaciones estándar) tienen más impacto en el valor final de similitud. Consideremos otro ejemplo para ilustrar este punto. Supongamos que queremos agrupar individuos mediante tres variables —actitudes hacia un producto, edad y renta. Ahora supongamos que medimos la actitud con una escala de siete puntos entre el agrado y el desagrado, la edad en años y la renta el dólares. Si dibujamos un gráfico tridimensional, la distancia entre los puntos (y sus similitudes) estaría basada casi completamente en las diferencias de renta. Las posibles diferencias de actitud van de 1 a 7, mientras que la renta tiene un rango de quizá mil veces mayor. Por tanto, gráficamente no seríamos capaces de ver ninguna diferencia en la dimensión asociada con la actitud. Por esta razón, el investigador debe tener cuidado con la ponderación implícita de las variables en función de su dispersión relativa, que sucede con las medidas de distancia.

Estandarización por variables

La forma más común de estandarización es la conversión de cada variable a unas puntuaciones estándar (también conocidas como puntuaciones Z) restando la media y dividiendo por la desviación típica de cada variable. Esta es una opción que tienen todos los programas informáticos y muchas veces incluso está incluida en el procedimiento de análisis *cluster*. Es una forma general de una

función de distancia normalizada que utiliza una medida de distancia Euclídea susceptible de transformación normalizada de los datos originales. Este proceso convierte cada puntuación de los datos originales en un valor estandarizado con una media de 0 y desviación estándar de 1. Esta transformación, a cambio, elimina el sesgo introducido por las diferencias en las mediciones de varios atributos o variables utilizadas en el análisis.

Los beneficios de la estandarización pueden verse en la última sección de la Tabla 9.4, en la que dos variables (probabilidad de compra y tiempo de visión) se han estandarizado antes de calcular las tres medidas de distancia. En primer lugar, es mucho más fácil comparar entre las variables en la medida en que están en la misma escala (una media 0 y desviación estándar 1). Los valores positivos están por encima de la media y los valores negativos están por debajo; la magnitud representa el número de desviaciones estándar del valor original a partir de la media. En segundo lugar, no existe diferencia entre los valores estandarizados cuando sólo cambia la escala. Por ejemplo, cuando el tiempo de visión en minutos y en segundos se estandariza, los valores son los mismos. Por tanto, al utilizar las variables estándar se eliminan verdaderamente los efectos debidos a las diferencias de escala no sólo entre las variables, sino también para la misma variable. Sin embargo, el investigador no debería aplicar sistemáticamente la estandarización sin considerar sus consecuencias. No hay razón para aceptar absolutamente una solución *cluster* utilizando variables estandarizadas frente a variables no estandarizadas. Si existe alguna relación «natural» reflejada en la escala de las variables, entonces la estandarización puede no ser apropiada. La decisión de estandarizar tiene impactos tanto conceptuales como empíricos y debería de hacerse siempre después de una cuidadosa consideración.

Estandarización por la observación

Hasta ahora hemos discutido la estandarización sólo para variables. Pero ¿qué pasa con la «estandarización» de encuestados o casos? ¿Por qué haríamos esto? Consideremos un simple ejemplo. Suponga que habíamos recogido de los encuestados un número de calificaciones sobre una escala de 10 puntos acerca de la importancia de varios atributos en sus decisiones de compra de un producto. Podríamos aplicar el análisis *cluster* y obtener conglomerados, pero una posibilidad muy distinta es que obtuviéramos conglomerados de gente que dice que algo era importante, y de gente que dice que algo tenía poca importancia, y quizá algunos conglomerados intermedios. Lo que estamos viendo son los **efectos de tipo de respuesta** en los conglomerados. Los efectos de tipo de respuesta son las pautas sistemáticas de respuesta a un conjunto de cuestiones, tales como los que siempre dicen sí (respuesta favorable para todas las cuestiones) o los que siempre dicen no (respuesta desfavorable para todas las cuestiones).

Si queremos identificar los grupos de acuerdo a su estilo de respuestas, entonces la estandarización no es apropiada. Pero en la mayoría de los casos lo que se desea es la importancia *relativa* de una variable a otra. En otras palabras, ¿es más importante el atributo 1 que el resto de los atributos, y pueden encontrarse en los conglomerados de encuestados pautas de importancia similar? En este caso, la estandarización por encuestado estadarizaría cada cuestión no por la media de la muestra sino por la puntuación media del encuestado. Esta **tipificación entre sujetos** o **tipificación centrada por filas** puede ser bastante efectiva al eliminar efectos de respuesta y es especialmente adecuada para muchas formas de datos de actitud [14]. Notemos que esto es similar a una medida de correlación en la señalización de la pauta para todas las variables, pero la proximidad de todos los casos todavía determina el valor de similitud.

Tercer paso: Supuestos del análisis cluster

El análisis *cluster*, como el análisis multidimensional (véase Capítulo 10), no es una técnica de inferencia estadística en la que se analizan los parámetros de una muestra en la medida en que puedan ser representativos de una población. Por el contrario, el análisis *cluster* es una metodología objetiva de cuantificación de las características estructurales de un conjunto de observaciones. Como tal, tiene fuertes propiedades matemáticas pero no fundamentos estadísticos. Las exigencias de nor-

malidad, linealidad y homocedasticidad que eran tan importantes en otras técnicas realmente tienen poco peso en el análisis *cluster*. El investigador debe centrarse, sin embargo, en otros dos asuntos críticos: la representatividad de la muestra y la multicolinealidad.

Representatividad de la muestra

Rara vez tiene el investigador un censo de la población con el fin de utilizarlo en el análisis *cluster*. Normalmente se obtiene una muestra de casos y los conglomerados se derivan en la esperanza de que representen la estructura de la población. El investigador debe, por tanto, confiar en que la muestra obtenida es verdaderamente representativa de la población. Como ya se ha mencionado, los atípicos pueden ser en realidad producto de una muestra escasa de grupos divergentes que, cuando se descartan, introducen sesgos en la estimación de la estructura. El investigador debe darse cuenta de que el análisis cluster sólo es tan bueno como la representatividad de la muestra. Por consiguiente, todos los esfuerzos deberían dirigirse a asegurar que la muestra es representativa y que los resultados son generalizables para la población a estudiar.

Impacto de la multicolinealidad

La **multicolinealidad** era materia de otras técnicas multivariantes a causa de la dificultad de discernir el «verdadero» impacto de las variables multicolineales. Pero en el análisis *cluster* el efecto es diferente porque aquellas variables que son multicolineales están implícitamente ponderadas con más fuerza. Comencemos con un ejemplo que ilustra el efecto de la multicolinealidad. Suponga que se está agrupando a los encuestados sobre 10 variables, todas afirmaciones de actitud hacia un servicio. Cuando se examina la multicolinealidad, vemos que existen realmente dos conjuntos de variables, el primero constituido de ocho afirmaciones y el segundo consistente en las dos restantes. Si nuestro intento es en realidad agrupar a los encuestados a partir de las dimensiones del producto (en este caso representadas por dos grupos de variables), entonces será un error utilizar las 10 variables iniciales. Dado que cada variable se pondera igualmente en el análisis *cluster*, la primera dimensión tendría como mucho cuatro veces más posibilidad (ocho ítems frente a dos) para afectar a la medida de similitud, y lo mismo ocurrirá con la segunda dimensión.

La multicolinealidad actúa como proceso de ponderación no aparente para el observador pero que sin embargo afecta al análisis. Por esta razón, se anima al investigador a examinar las variables utilizadas en el análisis *cluster* en busca de una multicolinealidad substancial y, si se encuentra, o bien reducir las variables al mismo número en cada conjunto o bien utilizar una de las distancias medidas, como la distancia de Mahalanobis, que compensa esta correlación. Hay un debate sobre el uso de las puntuaciones de factor en el análisis *cluster*, como cierta investigación ha mostrado que las variables que verdaderamente discriminan entre los grupos subyacentes que no están bien representados en la mayoría de las soluciones de factor. Por tanto, cuando se utilizan puntuaciones de factor, es bastante posible que se obtenga una mala representación de la verdadera estructura de los datos [13]. El investigador debe tratar tanto con la multicolinealidad como con la discriminabilidad de las variables para llegar a la mejor representación de la estructura.

Cuarto paso: Obtención de conglomerados y valoración del ajuste conjunto

Con las variables seleccionadas y la matriz de similitud calculada, comienza el proceso de partición (véase Figura 9.6). El investigador debe seleccionar en primer lugar el algoritmo de aglomeración utilizado en la formación de conglomerados y a continuación tomar la decisión del número de conglomerados que se van a formar. Ambas decisiones tienen implicaciones substanciales no sólo sobre los resultados que se obtendrán, sino también sobre la interpretación que se puede derivar de los resultados.

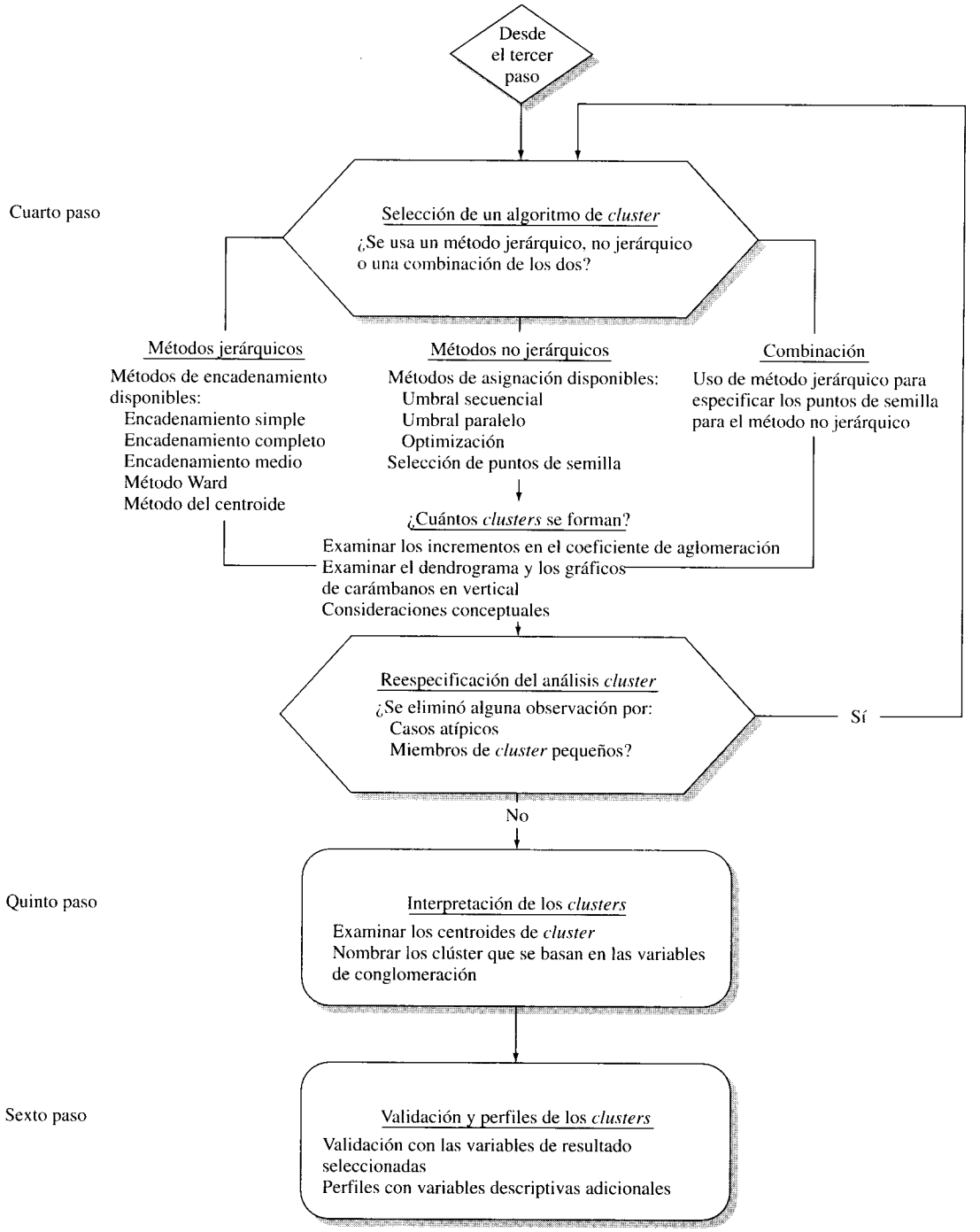


FIGURA 9.6. Pasos 4-6 del diagrama de decisión del análisis *cluster*

Algoritmo para la obtención de conglomerados

La primera cuestión importante que hay que contestar en la fase de partición es, ¿qué procedimiento debería utilizarse para colocar objetos similares en grupos o conglomerados? Esto es, ¿qué algoritmo de obtención de conglomerados o qué conjunto de reglas es más apropiado? No se trata de una pregunta sencilla puesto que existen cientos de programas informáticos que utilizan diferentes algoritmos, y siempre se están desarrollando más. El criterio esencial de todos los algoritmos

es, sin embargo, que intentan maximizar las diferencias entre los conglomerados relativa a la variación dentro de los conglomerados, tal y como se muestra en la Figura 9.7. La razón entre la variación entre conglomerados y la media de la variación dentro de los conglomerados es comparable (pero no idéntica) a la razón F del análisis de la varianza.

Los algoritmos de obtención de los conglomerados más utilizados pueden clasificarse en dos categorías generales: (1) jerárquicos y (2) no jerárquicos. Discutiremos en primer lugar las técnicas jerárquicas.

Procedimientos de obtención de conglomerados jerárquicos

Los **procedimientos jerárquicos** consisten en la construcción de una estructura en forma de árbol. Existen básicamente dos tipos de procedimientos de obtención de conglomerados jerárquicos —de aglomeración y divisivos. En los **métodos de aglomeración**, cada objeto u observación empieza dentro de su propio conglomerado. En etapas ulteriores, los dos conglomerados más cercanos (o individuos) se combinan en un nuevo conglomerado agregado, reduciendo así el número de conglomerados paso a paso. En algunos casos, un tercer individuo se une a los dos primeros en un conglomerado. En otros, dos grupos de individuos formados en un paso anterior pueden unirse en un nuevo conglomerado. Eventualmente, todos los individuos se agrupan en un único conglomerado; por esta razón, los procedimientos de aglomeración son denominados a veces como métodos de construcción.

Una característica importante de los procedimientos jerárquicos es que los resultados obtenidos en un paso previo siempre necesitan encajarse dentro de los resultados del siguiente paso, creando algo parecido a un árbol. Por ejemplo, una solución de seis conglomerados se obtiene uniendo dos de los conglomerados encontrados en el paso de siete conglomerados. Dado que los conglomerados se forman sólo por unión de los conglomerados existentes, se puede rastrear hasta su origen de simple observación cualquier miembro de un conglomerado. Se muestra este proceso en la Figura 9.8; la representación se denomina **dendrograma** o gráfico en forma de árbol. Otro método gráfico muy habitual es el **diagrama de carámbanos en vertical**.

Cuando el proceso de obtención de conglomerados procede en dirección opuesta al método de aglomeración, se denomina **método divisivo**. En los métodos divisivos, empezamos con un gran conglomerado que contiene todas las observaciones (objetos). En los pasos sucesivos, las observaciones que son más diferentes se dividen y se construyen conglomerados más pequeños. Este proceso continúa hasta que cada observación es un conglomerado en sí mismo. En la Figura 9.8, los

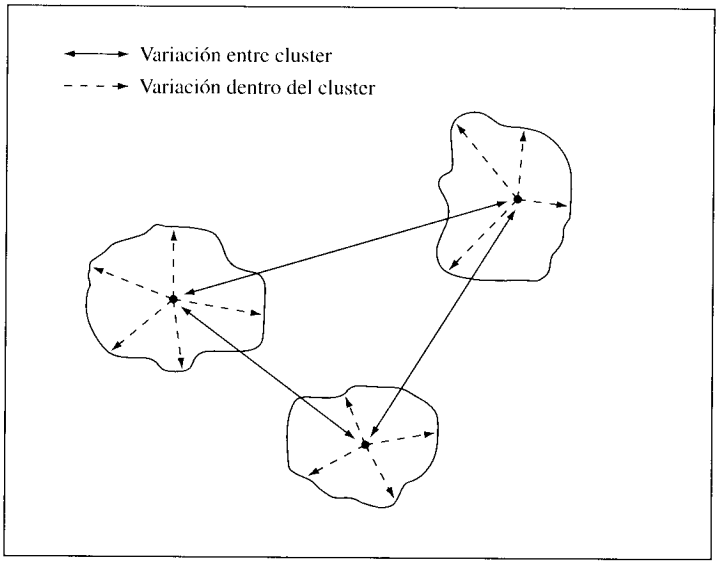


FIGURA 9.7. Diagrama de conglomerados que muestra la variación dentro y entre conglomerados.

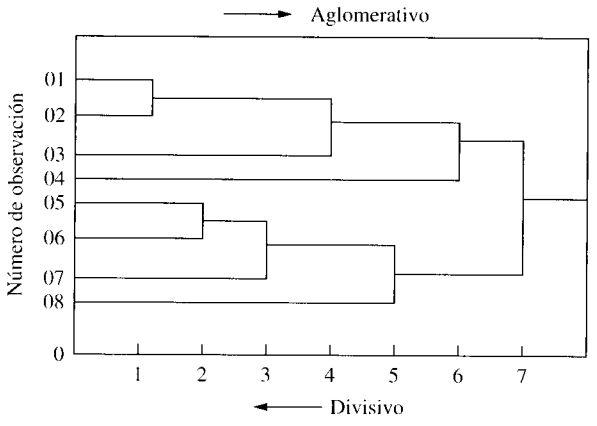


FIGURA 9.8. Dendrograma ilustrativo de la obtención de conglomerados jerárquicos.

métodos aglomerativos que van de izquierda a derecha y los métodos divisivos van de derecha a izquierda. Dado que los programas informáticos más habituales utilizan los métodos aglomerativos, y los métodos divisivos actúan como métodos aglomerativos al revés, nos centraremos en los métodos aglomerativos.

Son cinco los algoritmos más habituales utilizados para desarrollar conglomerados (1) método de encadenamiento simple, (2) encadenamiento completo, (3) encadenamiento medio, (4) método de Ward, y (5) método del centroide. Estos algoritmos difieren en cómo se calcula la distancia entre los conglomerados.

Encadenamiento simple El procedimiento del **encadenamiento simple** se basa en la distancia mínima. Encuentra los dos objetos separados por la distancia más corta y las coloca en el primer conglomerado. A continuación se encuentra la distancia más corta, y o bien un tercer objeto se une a los dos primeros para formar un conglomerado o se forma un nuevo conglomerado de dos miembros. El proceso continúa hasta que todos los objetos se encuentran en un conglomerado. Este procedimiento también se ha denominado como el enfoque del vecino más cercano.

La distancia entre dos conglomerados cualquiera es la distancia más corta desde cualquier punto en un conglomerado a cualquier punto en el otro. Dos conglomerados se fusionan en cualquier nivel por el vínculo más corto o más fuerte entre ellos. Fue una regla aplicada en el ejemplo del principio de este capítulo. Los problemas se producen, sin embargo, cuando los conglomerados están mal definidos. En tales casos, los procedimientos de encadenamientos simples pueden formar largas y sinuosas cadenas, y eventualmente todos los individuos pueden situarse en una ca-

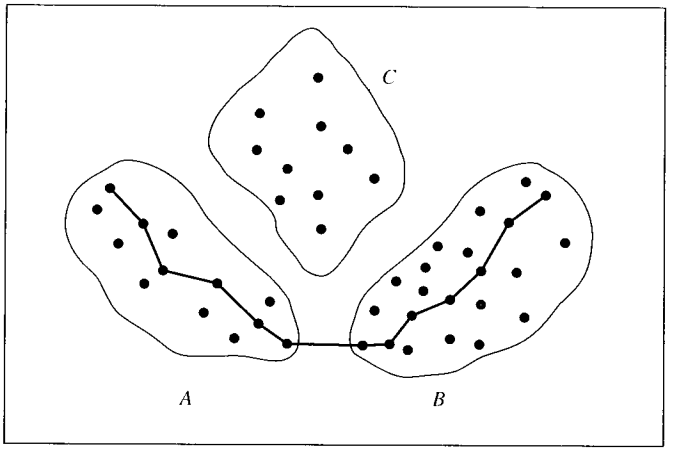


FIGURA 9.9. Ejemplo de encadenamiento simple que une los conglomerados diferentes A y B.

dena. Los individuos que se encuentran en los límites opuestos de una cadena pueden ser muy diferentes.

Un ejemplo de este acuerdo se muestra en la Figura 9.9. Se van a unir tres conglomerados (A, B y C). El algoritmo de encadenamiento simple, centrándose sólo en los puntos más cercanos de cada conglomerado, uniría los conglomerados A y B debido a su distancia más reducida en los extremos de los conglomerados. Uniendo los conglomerados A y B se crea un conglomerado que rodea al conglomerado C. Pero si buscamos la homogeneidad dentro del conglomerado, sería mucho mejor juntar el conglomerado C con A o B. Esta es la principal desventaja del algoritmo de encadenamiento simple.

Encadenamiento completo El procedimiento de **encadenamiento completo** es parecido al del encadenamiento simple excepto en que el criterio de aglomeración se basa en la distancia máxima. Por esta razón, a veces se le denomina como aproximación del vecino más lejano o método del diámetro. La distancia máxima entre individuos de cada conglomerado representa la esfera más reducida (diámetro mínimo) que puede incluir todos los objetos en ambos conglomerados. A este método se le denomina encadenamiento completo porque todos los objetos de un conglomerado se vinculan con el resto a alguna distancia máxima o por la mínima similitud. Podemos decir que la similitud dentro del grupo es igual al diámetro del grupo. Esta técnica elimina el problema identificado para el encadenamiento simple.

La Figura 9.10 muestra cómo las distancias más cortas (encadenamiento simple) o las más largas (encadenamiento completo) representan la similitud entre grupos. Ambas medidas reflejan sólo un aspecto de los datos. El uso de la distancia más corta refleja sólo un único par de objetos (los más cercanos) y el encadenamiento completo también refleja un único par, esta vez los dos más lejanos. Por tanto, es útil ver las medidas como reflejo de la similitud del par de objetos más parecido o del par menos parecido.

Encadenamiento medio El método del **encadenamiento medio** comienza igual que los métodos de encadenamiento simple o completo, pero el criterio de aglomeración es la distancia media de todos los individuos de un conglomerado con todos los individuos de otro. Tales técnicas no dependen de los valores extremos, como se hace en el encadenamiento simple o completo y la partición se basa en todos los miembros de los conglomerados en lugar de un par único de miembros extremos. El enfoque del encadenamiento medio tiende a combinar los conglomerados con variaciones reducidas dentro del conglomerado. También tiende a estar sesgado hacia la producción de conglomerados con aproximadamente la misma varianza.

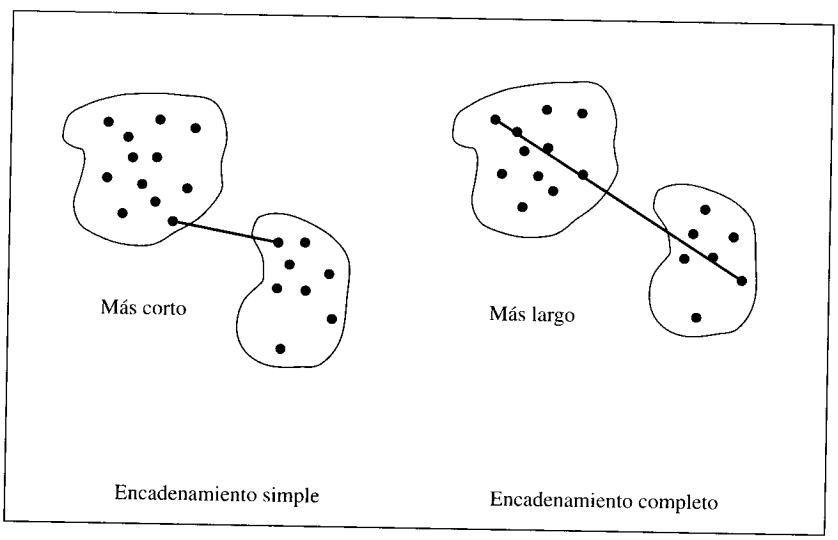


FIGURA 9.10. Comparación de las medidas de distancia del encadenamiento simple y completo.

Método de Ward En el **método de Ward**, la distancia entre dos conglomerados es la suma de los cuadrados entre dos conglomerados sumados para todas las variables. En cada paso del procedimiento de aglomeración, se minimiza la suma de los cuadrados dentro del conglomerado para todas las particiones (el conjunto completo de conglomerados disjuntos o separados) obtenida mediante la combinación de dos conglomerados en un paso previo. Este procedimiento tiende a combinar los conglomerados con un número reducido de observaciones. También está sesgado hacia la producción de conglomerados con aproximadamente el mismo número de observaciones.

Método del centroide En el **método del centroide** la distancia entre los dos conglomerados es la distancia (normalmente Euclídea simple o cuadrada) entre sus centroides. Los **centroides de los grupos** son los valores medios de las observaciones de las variables en el valor teórico del conglomerado. En este método, cada vez que se agrupa a los individuos, se calcula un nuevo centroide. Los centroides de los grupos cambian a medida que se fusionan conglomerados. En otras palabras, existe un cambio en un centroide de un grupo cada vez que un nuevo individuo o grupo de individuos se añade al conglomerado existente. Estos métodos son más populares entre los biólogos pero pueden producir resultados desordenados y a menudo confusos. La confusión se produce a causa de los cambios, esto es, casos donde la distancia entre los centroides de un par puede ser menor que la distancia entre los centroides de otro par fusionado en una combinación anterior. La ventaja de este método es que se ve menos afectada por los atípicos que otros métodos jerárquicos.

Procedimientos no jerárquicos de obtención de conglomerados

En contraste con los métodos jerárquicos, los **procedimientos no jerárquicos** no implican los procesos de construcción de árboles. En su lugar, asignan los objetos a conglomerados una vez que el número de conglomerados a formar está especificado. Por tanto, la solución de seis conglomerados no es sólo una combinación de dos conglomerados desde una solución de siete conglomerados, sino que se basa sólo en la búsqueda de la mejor solución de seis conglomerados. En un ejemplo simple, el proceso opera de la siguiente forma. El primer paso es seleccionar una **semilla de conglomerado** como centro de conglomerado inicial, y todos los objetos (individuos) dentro de una distancia umbral previamente especificada se incluyen dentro del conglomerado resultante. Entonces se selecciona otra semilla de conglomerado y la asignación continúa hasta que todos los objetos están asignados. Los objetos pueden entonces asignarse si están cercanos a otro conglomerado que no sea el original. Existen diferentes aproximaciones para seleccionar las semillas de conglomerado y asignar objetos que discutiremos en la siguiente sección. Los procedimientos de aglomeración no jerarquizados se denominan frecuentemente como aglomeración de K-medias, y normalmente utilizan uno de las siguientes tres aproximaciones para asignar las observaciones individuales de uno de los conglomerados [5].

Umbral secuencial El **método del umbral secuencial** empieza seleccionando una semilla de conglomerado e incluye todos los objetos que caen dentro de una distancia previamente especificada. Cuando todos los objetos dentro de la distancia están incluidos, se selecciona una segunda semilla de conglomerado y se incluyen todos los objetos dentro de la distancia previamente especificada. A continuación se selecciona una tercera semilla, y el proceso continúa como se ha descrito. Cuando un objeto se incluye en un conglomerado con una semilla, no se considera a efectos de ulteriores semillas.

Umbral paralelo Como contraste, el **método del umbral paralelo** selecciona varias semillas de conglomerado simultáneamente al principio y asigna objetos dentro de la distancia umbral hasta la semilla más cercana. A medida que el proceso avanza, se puede ajustar las distancias umbral para incluir más o menos objetos en los conglomerados. También, en algunas variantes de este método, los objetos permanecen fuera de los conglomerados si están fuera de la distancia previamente especificada desde cualquiera de las semillas de conglomerado.

Optimización El tercer método, denominado como **procedimiento de optimización**, es parecido a los otros dos procedimientos no jerárquicos excepto en que permite la reubicación de los objetos. Si, en el curso de la asignación de los objetos, un objeto se acerca más a otro conglomerado que no es el que tiene asignado en este momento, entonces un procedimiento de optimización cambia el objeto al conglomerado más parecido (cercano).

Selección de puntos de semilla Los procedimientos no jerárquicos se encuentran disponibles en varios programas informáticos, incluyendo los principales programas estadísticos. El procedimiento del umbral secuencial (por ejemplo, el programa FASTCLUS en SAS) es un ejemplo de programa de formación de conglomerados no jerarquizado diseñado para conjuntos con gran cantidad de datos. Una vez que el investigador especifica el número máximo de conglomerados permitidos, el procedimiento comienza con la selección de semillas de conglomerados, que se utilizan como conjeturas iniciales de las medias de los conglomerados. La primera semilla es la primera observación del conjunto de datos sin valores perdidos. La segunda semilla es la siguiente observación completa (sin datos perdidos) que se separa de la primera semilla mediante una distancia mínima especificada. La opción por defecto es una distancia mínima de cero. Una vez que se han seleccionado todas las semilla, el programa asigna cada observación al conglomerado con la semillas más próxima. El investigador puede especificar que los conglomerados de semillas se revisen (actualicen) mediante el cálculo de medias de los conglomerados de semillas cada vez que se asigna una observación. Como contraste, los métodos del umbral paralelo (por ejemplo, QUICKCLUSTER en SPSS) establecen los puntos de semilla como puntos aportados por el usuario o seleccionado aleatoriamente de las observaciones.

El principal problema a que se enfrentan todos los métodos de formación de conglomerados no jerárquicos es cómo seleccionar las semillas de conglomerado. Por ejemplo, con una opción de umbral secuencial, los resultados del conglomerado inicial y probablemente del final dependerán del orden de las observaciones en el conjunto de datos y arrastrar el orden de los datos es como afectar a los resultados. La especificación de las semillas de conglomerado iniciales, como se hace en el procedimiento de umbral secuencial, puede reducir este problema. Pero incluso la selección aleatoria de las semillas de conglomerado producirá diferentes resultados para cada conjunto de puntos de semilla aleatorios. Por tanto, el investigador debe ser consciente del impacto del proceso de selección de las semillas de conglomerados en los resultados finales.

¿Deben utilizarse los métodos jerárquicos o no jerárquicos?

No puede darse una respuesta definitiva a esta cuestión por dos razones. En primer lugar, el problema a investigar en ese momento puede sugerir un método u otro. En segundo lugar, lo que aprendemos con la continua aplicación a un contexto particular puede sugerir un método u otro como el más aconsejable para ese contexto.

Pros y contras de los métodos jerárquicos En el pasado, las técnicas jerárquicas de formación de conglomerados eran las más populares, siendo el método de Ward y el encadenamiento medio probablemente los mejores disponibles [8]. Los procedimientos jerárquicos tienen la ventaja de ser más rápidos y llevar menos tiempo de cálculo. Pero con el poder de cálculo de hoy en día, incluso los computadores personales pueden manejar grandes conjuntos de datos fácilmente. Los métodos jerárquicos pueden dar una idea equivocada, sin embargo, porque combinaciones iniciales indeseables pueden persistir a lo largo del análisis y llevar a resultados artificiales. De interés específico es el impacto substancial de los atípicos sobre los métodos jerárquicos, particularmente con el método del encadenamiento completo. Para reducir esta posibilidad, el investigador puede querer realizar el análisis *cluster* de los datos repetidas veces, eliminando los atípicos o las observaciones problemáticas. La destrucción de casos, sin embargo, incluso aquellos que no sean atípicos, puede muchas veces distorsionar la solución. Por tanto, el investigador debe tener un cuidado extremo en la destrucción de las observaciones por la razón que sea.

También, aunque los cálculos de los procesos de formación de conglomerados son relativamente rápidos, los métodos jerárquicos no son susceptibles de analizar muestras muy grandes. A medida que aumenta el tamaño de la muestra, los requisitos de almacenamiento de datos aumentan enormemente. Por ejemplo, una muestra de 400 casos exige el almacenamiento de aproximadamente 80.000 similitudes, que se incrementa a 125.000 para una muestra de 500. Incluso con los avances tecnológicos actuales, problemas de este calibre exceden la capacidad de la mayoría de los computadores actuales, limitando por tanto la aplicación en muchos casos. El investigador puede considerar una muestra aleatoria de las observaciones originales para reducir el tamaño de la muestra pero debe cuestionarse ahora la representatividad de la muestra tomada de la muestra original.

Aparición de los métodos no jerárquicos Los métodos no jerarquizados han ganado una creciente aceptación y se aplican cada vez más. Su uso, sin embargo, depende de la capacidad del investigador para seleccionar los puntos de semilla de acuerdo a bases prácticas, objetivas y teóricas. En estos casos, los métodos no jerárquicos tienen varias ventajas sobre las técnicas jerárquicas. Los resultados son menos susceptibles a los datos atípicos, a la medida de distancia utilizada y a la inclusión de variables irrelevantes o inapropiadas. Estos beneficios se obtienen, sin embargo, sólo con el uso de puntos de semillas no aleatorios (es decir, especificados); por tanto, el uso de técnicas no jerárquicas con puntos de semilla aleatorios es notablemente inferior a las técnicas jerárquicas. Incluso comenzando con una solución no aleatoria no se garantiza una formación de conglomerados óptima de observaciones. De hecho, en muchos casos, el investigador obtendrá una solución final diferente para cada conjunto de puntos de semilla especificados. ¿Cómo va a seleccionar el investigador la respuesta «correcta»? Sólo mediante el análisis y la validación puede el investigador seleccionar lo que se considera la «mejor» representación de la estructura, teniendo presentes las muchas alternativas que pueden considerarse aceptables.

Una combinación de ambos métodos Otra aproximación es utilizar *ambos* métodos (jerárquico y no jerárquico) para obtener los beneficios de cada uno [8]. En primer lugar, una técnica jerárquica puede establecer el número de conglomerados, los perfiles de los centros de conglomerados y la identificación de cualquier atípico obvio. Una vez que se han eliminado los atípicos, las observaciones restantes pueden ser agrupadas mediante un método no jerárquico con los centros de conglomerados desde los resultados jerárquicos como los puntos de semillas iniciales. De esta forma, las ventajas de los métodos jerárquicos se complementan con la capacidad de los métodos no jerárquicos para «ajustar» los resultados permitiendo el cambio de pertenencia a un conglomerado.

¿Cuántos grupos deben formarse?

Quizá el asunto más desconcertante para el investigador que utiliza el análisis de conglomerados es la determinación del número final de conglomerados a formar (también conocida como **regla de parada**). Desafortunadamente no existe un procedimiento objetivo o estándar. Dado que no se utiliza un criterio estadístico interno para la inferencia, tal como los test de significación estadística de otros métodos multivariantes, los investigadores han desarrollado varios criterios y líneas a seguir para aproximarse al problema. La principal conclusión es que existen procedimientos *ad hoc* que deben ser calculados por el investigador, lo que muchas veces implica procedimientos francamente complejos [1, 9]. Una clase de reglas de parada que es relativamente simple examina alguna medida de similitud o distancia entre los conglomerados a cada paso sucesivo, donde la solución cluster se define cuando la medida de similitud excede a un valor especificado o cuando los valores sucesivos entre los pasos dan un salto súbito. Un ejemplo sencillo ya se utilizó en el ejemplo del principio del capítulo, que buscaba grandes aumentos en la media de la distancia dentro del conglomerado. Cuando se produce un gran aumento, el investigador selecciona la solución *cluster* previa en la lógica de que su combinación provocó la substancial reducción en su similitud. Se ha mostrado que esta regla de parada ofrece decisiones francamente precisas en los estu-

dios empíricos [9]. Una segunda clase de reglas de parada intentan aplicar alguna forma de regla estadística o adaptar un test estadístico, tal como las correlaciones «point-biserial/tau» o el ratio de verosimilitud. Aunque alguno de estos criterios —como el criterio cúbico de elaboración de conglomerados (CCC) que se encuentra en el SAS— han mostrado tener un éxito notable, muchos parecen demasiado complejos para la mejora que ofrecen sobre muestras simples. Se ha propuesto cierto número de procedimientos específicos, pero no se ha encontrado ninguno que sea mejor en todas las situaciones.

También, el investigador debería complementar el juicio estrictamente empírico con cualquier conceptualización de las relaciones teóricas que pueda sugerir un número natural de conglomerados. Se puede empezar este proceso especificando algún criterio basándose en consideraciones prácticas, como decir, «mis resultados serán más manejables y más fáciles de comunicar si tengo de tres a seis conglomerados», y a continuación resolver para este número de conglomerados y seleccionar la mejor alternativa después de evaluar todas ellas. En el análisis final, sin embargo, probablemente sea mejor calcular varias soluciones *cluster* diferentes (es decir, dos, tres, cuatro) y después decidir entre las soluciones alternativas utilizando criterios a priori, juicios prácticos, sentido común o fundamentos teóricos. Las soluciones *cluster* se verán mejoradas mediante la restricción de la solución de acuerdo con los aspectos conceptuales del problema.

¿Debe volverse a especificar el análisis *cluster*?

Cuando se identifica una solución aceptable en el análisis *cluster*, el investigador debería examinar la estructura fundamental representada en los conglomerados definidos. De interés particular son los tamaños claramente disparatados de los conglomerados o conglomerados de sólo una o dos observaciones. Los investigadores deben examinar los tamaños de conglomerados muy variables desde una perspectiva conceptual, comparando los resultados actuales con las expectativas formadas en los objetivos de investigación. Más problemáticos son los conglomerados de un único miembro, que pueden ser atípicos no detectados en análisis anteriores. Si aparece un conglomerado de un único miembro (o uno de un tamaño muy pequeño comparado con otros conglomerados), el investigador debe decidir si representa un componente estructural válido de la muestra o si debería ser eliminado como no representativo. Si se destruye cualquier observación, concretamente cuando se emplean soluciones jerárquicas, el investigador debería repetir el análisis cluster y empezar de nuevo el proceso de definición de los conglomerados.

Quinto paso: Interpretación de los conglomerados

El paso de la interpretación implica el examen de cada conglomerado en términos del valor teórico del conglomerado o asignar una etiqueta precisa que describa la naturaleza de los conglomerados. Para clarificar este proceso, vamos a referir un ejemplo de refrescos normales y «lights». Supongamos que se desarrolla una escala de actitud que consista en afirmaciones en relación con el consumo de refrescos, tales como «los refrescos light tienen un sabor áspero», «los refrescos light tienen mucho sabor», «los refrescos light son más saludables», etcétera. Además, supongamos que se obtienen también los datos demográficos y de consumo de refrescos.

Cuando se comienza el proceso de interpretación, una medida utilizada frecuentemente es el centroide del conglomerado. Si el procedimiento de aglomeración se realizó sobre los datos tal y como se obtuvieron, esto sería una descripción lógica. Si los datos se estandarizaron o si el análisis *cluster* se realizó utilizando el análisis factorial (de componentes principales), el investigador tendría que retroceder a las puntuaciones dadas por los encuestados de las variables originales y calcular los perfiles medios utilizando estos datos.

Continuando con nuestro ejemplo de los refrescos, en este paso examinamos los perfiles de las puntuaciones medias sobre las afirmaciones acerca de las actitudes para cada grupo y asignamos una etiqueta descriptiva para cada conglomerado. Muchas veces el análisis discriminante se apli-

ca para generar puntuaciones de perfiles, pero debemos recordar que las diferencias estadísticamente significativas no indicarían una solución «óptima» porque se esperan diferencias estadísticas, dados los objetivos del análisis cluster. El examen de los perfiles permite una descripción más rica de cada conglomerado. Por ejemplo, dos de los conglomerados pueden tener actitudes favorables sobre los refrescos «light» y el tercer conglomerado actitudes negativas. Además, de los dos conglomerados más favorables, uno puede exhibir actitudes favorables sólo hacia los refrescos «light», mientras que los otros pueden desplegar actitudes favoritas tanto hacia los refrescos normales como los «light». Desde este procedimiento analítico, uno evaluaría las actitudes de cada conglomerado y desarrollaría interpretaciones sustantivas para facilitar su clasificación. Por ejemplo, un conglomerado puede calificarse como «consciente de la salud y las calorías» mientras que otro puede calificarse como «goloso compulsivo».

Los perfiles y la interpretación de los conglomerados, sin embargo, consiguen algo más que una descripción. En primer lugar, proporcionan un medio de evaluar la correspondencia de los conglomerados derivados de aquellos propuestos por una teoría a priori o por la experiencia práctica. Si se utiliza de forma confirmatoria, los perfiles del análisis *cluster* ofrecen un medio directo de evaluación de la correspondencia. En segundo lugar, los perfiles de los conglomerados ofrecen una vía para realizar evaluaciones de significación práctica. El investigador puede exigir que existan diferencias substanciales en un conjunto de variables de elaboración de conglomerados y que las soluciones *cluster* aumenten hasta que surjan tales diferencias. Al evaluar tanto su correspondencia o significación práctica, el investigador compara los conglomerados derivados con una tipología preconcebida.

Sexto paso: Validación y perfil de los grupos

Dada la naturaleza de alguna forma subjetiva del análisis *cluster* sobre la selección de una solución *cluster* «óptima», el investigador debería tener mucho cuidado en la validación y asegurarse la relevancia práctica de la solución *cluster* definitiva. Aunque no existe un método único para asegurar la validez y la relevancia práctica, se han propuesto diferentes aproximaciones para ofrecer cierta base a la evaluación realizada por el investigador.

Validación de la solución *cluster*

La validación incluye los intentos del investigador por asegurar que la solución *cluster* es representativa de la población general y por tanto generalizable a otros objetos y estable en el tiempo. La aproximación más directa en este sentido es realizar análisis *cluster* para muestras distintas. Esta aproximación, sin embargo, a menudo no es práctica debido a las restricciones de tiempo o de costes o a la no disponibilidad de objetos (particularmente consumidores) para múltiples análisis *cluster*. En estos casos, una aproximación común es escindir la muestra en dos grupos. Cada conglomerado se analiza por separado y se comparan después los resultados. Otras aproximaciones incluyen (1) una forma modificada de escisión de muestra por la cual se emplean los centros de conglomerados obtenidos desde una solución *cluster* para definir conglomerados a partir de otras observaciones para comparar después los resultados [7], y (2) una forma directa de validación cruzada [12].

El investigador también puede intentar establecer alguna forma de **criterio** o **validez predictiva**. Para hacerlo, el investigador selecciona una variable o variables *no utilizadas para formar los conglomerados* pero que se sabe que cambian a lo largo de los conglomerados. En nuestro ejemplo, podemos conocer de la investigación pasada las actitudes hacia los refrescos «light» por grupos de edad. Por tanto, podemos contrastar estadísticamente las diferencias de edad entre aquellos conglomerados que sean favorables a los refrescos «light» y aquellos que no lo son. Las variables utilizadas para evaluar la validez predictiva deberían tener un fuerte apoyo teórico o práctico en la medida en que se conviertan en el punto de referencia de selección entre las soluciones de conglomerado.

Perfiles de la solución *cluster*

El paso de los perfiles implica la descripción de las características de cada conglomerado para explicar en qué medida pueden diferir en dimensiones relevantes, lo que implica típicamente el uso del análisis discriminante (véase Capítulo 5). El procedimiento comienza una vez que se han identificado los conglomerados. El investigador utiliza *datos no previamente incluidos* en el procedimiento *cluster* para perfilar las características de cada conglomerado. Esto datos son normalmente características demográficas, perfiles psicográficos, pautas de consumo, etcétera. Aunque puede no existir una razón teórica para que difieran entre los conglomerados, tal como requerir la evaluación de la validez predictiva, tienen que tener al menos importancia práctica. Utilizando el análisis discriminante, el investigador compara los perfiles de las puntuaciones medias para los conglomerados. La variable categórica dependiente es el conjunto de conglomerados previamente identificado y las variables independientes son las demográficas, psicográficas, etcétera. Desde este análisis, suponiendo la relevancia estadística, el investigador podría concluir, por ejemplo, que el conglomerado de «consciente de la salud y la calorías» de nuestro ejemplo previo consiste en profesionales de rentas altas y educación superior que son consumidores moderados de refrescos. En resumen, el análisis de perfil se centra en la descripción no de lo que directamente se determinan los conglomerados sino de las características de los conglomerados una vez que se han identificado. Además, se hace hincapié en las características que difieren significativamente entre los conglomerados y aquellas que podrían predecir la pertenencia a un conglomerado particular.

Resumen del proceso de decisión

El análisis *cluster* equipa a los investigadores con un método objetivo y empírico para desarrollar una de las tareas más inherentemente humanas —la clasificación. Sea a efectos de simplificación, exploración o confirmación, el análisis *cluster* es una potente herramienta analítica que tiene un amplio rango de aplicaciones. Pero con esta técnica viene también la responsabilidad por parte del investigador de aplicar los principios subyacentes apropiadamente. Como se mencionó en la introducción al capítulo, el análisis *cluster* tiene muchos inconvenientes, que hacen que incluso el investigador experimentado lo aplique con precaución. Pero cuando se utiliza apropiadamente, tiene el potencial de revelar estructuras dentro de los datos que no podían descubrirse de otra forma. Por tanto, esta poderosa técnica se dirige a satisfacer una necesidad fundamental de los investigadores en todos los campos, aunque puede aplicarse con el conocimiento de que se puede tanto abusar de ella como utilizarse sabiamente.

Un ejemplo ilustrativo

Para ilustrar la aplicación de las técnicas de análisis *cluster*, volvamos a la base de datos de HATCO. La siete percepciones de HATCO proporcionan una base para ilustrar uno de los usos más comunes del análisis *cluster* —la formación de segmentos de consumidores. En nuestro ejemplo, seguimos el proceso de construcción de modelos, comenzando con la fijación de objetivos, siguiendo con los supuestos de diseño de investigación y finalmente dividiendo a los encuestados en conglomerados y la interpretación y validación de resultados. Las siguientes secciones detallan estos procedimientos a través de cada una de las etapas.

Primer paso: Objetivos del análisis *cluster*

Empezamos analizando mediante conglomerados las calificaciones de los clientes de HATCO acerca de en qué medida HATCO cumple los siete atributos (X_1 hasta X_7). Nuestro objetivo es segmentar

los objetos (clientes) en grupos con percepciones similares de HATCO. Una vez que se ha identificado, HATCO puede formular estrategias con diferentes atractivos para distintos grupos —la base exigida para la segmentación de mercados. Un asunto importante es que los siete atributos utilizados para formar los conglomerados sean adecuados en alcance y detalle. De los ejemplos de otros capítulos con diversas técnicas multivariantes, hemos encontrado que estas variables tienen suficiente poder predictivo como para justificar su uso como base para la segmentación.

Segundo paso: Diseño de investigación del análisis *cluster*

El primer paso es identificar cualquier atípico de la muestra antes de empezar con las particiones. La muestra de 100 observaciones ha sido examinada buscando atípicos y se ha encontrado que no tiene candidatos firmes para la eliminación en relación con los supuestos del Capítulo 2 y la discusión sobre la detección de atípicos presentada en este capítulo. Pero el investigador debería también examinar las soluciones *cluster* en etapas posteriores para evaluar si han surgido atípicos durante el proceso de aglomeración. El siguiente paso implica la elección de una medida de similitud. Dado que el conjunto de siete variables es métrico, se escoge la distancia Euclídea al cuadrado como medida de similitud. Si se considera que la multicolinealidad es substancial o que tiene el efecto de desigualar las ponderaciones de las variables, entonces la distancia de Mahalanobis (D^2) sería la más apropiada. Además, no se emplean medidas de corrección a causa de que la derivación de segmentos debería considerar la magnitud de las percepciones (favorables frente a desfavorables) así como la pauta. Esto se hace mejor con una medida de distancia de similitud. Finalmente, no se utiliza ninguna forma de estandarización. La estandarización de variables no se considera porque todas las variables están en la misma escala, y la estandarización dentro del caso no es apropiada porque la magnitud de las percepciones es un elemento importante en el objetivo de segmentación.

Tercer paso: Supuestos del análisis *cluster*

A efectos ilustrativos, la muestra se considera representativa de los clientes de HATCO. Queda por resolver el impacto de la multicolinealidad de las ponderaciones implícitas de los resultados. El análisis de la multicolinealidad detallado en los Capítulos 3 y 4 identifica sólo niveles mínimos que no deberían influir en el análisis *cluster* de forma substancial.

Cuarto paso: Obtención de grupos y valoración del ajuste conjunto

En nuestro ejemplo, seguimos la aproximación que emplea una combinación de métodos jerárquicos y no jerárquicos. El primer paso en el paso de partición utilizada el procedimiento jerárquico para identificar el número apropiado de conglomerados. A continuación, en el Paso 2 utilizamos procedimientos no jerárquicos para «ajustar» los resultados aún más mediante la utilización de resultados jerárquicos como base para la generación de puntos de semilla. En este análisis se utilizan los procedimientos jerárquicos y no jerárquicos de SPSS, aunque se pueden obtener resultados comparables con cualquiera de los programas de obtención de conglomerados.

Paso 1: *Análisis* cluster *jerárquico*

La primera cuestión a plantearse es, ¿qué algoritmo de obtención de conglomerados deberíamos utilizar? Se escoge el método de Ward para minimizar las diferencias dentro del conglomerado y evitar problemas con el «encadenamiento» de las observaciones encontradas en el método de encadenamiento simple. La Tabla 9.5 contiene los resultados del análisis *cluster*, incluyendo los casos que se han combinado en cada paso del proceso y el coeficiente de aglomeración. El coeficiente de aglomeración (cuarta columna) es la suma de los cuadrados dentro del conglomerado. Para otros

métodos de encadenamiento, el coeficiente de aglomeración es la distancia Euclídea entre los dos casos de conglomerados que se están considerando.

Selección de una solución cluster ¿Cuántos conglomerados debería haber? Dado que los datos comprenden perfiles de los clientes de HATCO y lo que nos interesa es identificar los tipos o perfiles de estos clientes que pueden formar las bases de diferentes estrategias, un número manejable de conglomerados estaría en el rango de dos a cinco. El investigador debe ahora seleccionar la solución *cluster* final de estas posibles soluciones *cluster*.

Este coeficiente de aglomeración es particularmente susceptible de ser utilizado en una regla de parada que evalúa los cambios en el coeficiente a cada nivel del proceso jerárquico. Los coeficientes pequeños indican que están surgiendo conglomerados claramente homogéneos. Unir dos conglomerados muy diferentes produce un coeficiente muy grande o un gran porcentaje de cambio en el coeficiente. El investigador busca grandes aumentos del valor, similar al contraste de caída del análisis factorial. Se ha mostrado que este test es un algoritmo bastante preciso, aunque tiene la tendencia a indicar muy pocos conglomerados [9]. Otra regla de parada habitual es el criterio de obtención de conglomerados cúbico (CCC), popularizado por SAS, que no se encuentra en SPSS. Aunque el CCC ha tenido bastante éxito en la identificación del número correcto de conglomerado, tiene la tendencia a seleccionar las soluciones que tienen muchos conglomerados [9]. El investigador debería considerar un rango de soluciones *cluster*, que hace más hincapié en el examen de las tendencias de la regla particular de parada del análisis.

Los coeficientes de aglomeración en la obtención de conglomerados, muestran grandes aumentos que van desde los cuatro a los tres conglomerados (523,0 – 446,3 = 76,6), de tres a dos (615,0 – 523,0 = 92,0) y de dos a uno (994,8 – 615,0 = 379,8). Para ayudar a identificar grandes aumentos relativos en la homogeneidad de los conglomerados, calculamos el porcentaje de cambio del coeficiente de obtención de conglomerados de 10 a 2 conglomerados (véase Tabla 9.6). El mayor aumento del porcentaje se produce al ir de dos a un conglomerado, y el siguiente cambio notable en el aumento del porcentaje se produce al combinar cuatro conglomerados en tres. Por tanto, a efectos ilustrativos así como en la evaluación de la significación práctica de la solución de cuatro conglomerados, se examinan las soluciones de dos a cuatro conglomerados.

Identificación de atípicos de las soluciones cluster El dendrograma (véase Figura 9.11) y el esquema de aglomeración (véase Tabla 9.5) proporcionan un medio de identificación de atípicos de la muestra. El dendrograma permite una inspección visual de los atípicos, donde un atípico sería una «rama» que no se unió hasta muy tarde. El investigador puede también identificar fácilmente conglomerados pequeños, en la medida en que exhiben grandes «ramas» únicamente para un número reducido de observaciones. En el esquema de aglomeración, el investigador puede averiguar la presencia de conglomerados de un único componente con facilidad mediante diversos programas informáticos. El ejemplo de la Tabla 9.5 del SPSS muestra en el lado izquierdo del coeficiente de aglomeración los conglomerados que se están combinando. En las columnas de la mano derecha, se anotan los pasos en que cada conglomerado se va formando. Una observación que nunca se ha unido a un conglomerado tiene un nivel 0. En los primeros 38 pasos, las observaciones únicas se van uniendo. Sólo en el paso 39 el análisis *cluster* consigue unir un conglomerado formado en otro paso. Podemos utilizar esta información también para identificar observaciones únicas que se unen tardíamente al proceso de elaboración de conglomerados —potenciales atípicos. Mirando hacia el Paso 99 de la Tabla 9.5, vemos que en el Paso 94 (seis conglomerados) se unió un conglomerado formado en el Paso 2. Esto significa que si seleccionamos una solución de siete conglomerados, uno de los conglomerados tendría sólo dos observaciones. Podemos ver también que el conglomerado de miembro único se unió en el Paso 79. Por tanto, si el análisis se confina en un número reducido de conglomerados (digamos, 10 o menos), entonces el investigador sólo tiene un problema potencial (el conglomerado de dos miembros) que tratar. En este caso, la selección de menos de siete conglomerados elimina la necesidad de cualquier re-especificación ulterior del análisis *cluster*.

Perfil de las soluciones de dos y cuatro conglomerados Sin necesidad de re-especificar, el siguiente paso es conseguir los perfiles de los conglomerados en ambas soluciones para

TABLA 9.6. Análisis del coeficiente de aglomeración para el análisis *cluster* jerárquico

Número de conglomerado	*Coeficiente de aglomeración*	*Cambio porcentual en el coeficiente del nivel siguiente*
10	258,7	8,8
9	281,4	8,4
8	305,0	9,2
7	333,1	9,6
6	364,9	9,1
5	398,1	12,1
4	446,3	17,2
3	523,0	17,6
2	615,0	61,8
1	994,8	—

apoyar la selección de la solución *cluster* final. La Tabla 9.7 contiene los perfiles de las variables de la obtención de conglomerados tanto para las soluciones de dos y cuatro conglomerados. El interés en este paso no es ofrecer una interpretación del conglomerado, sino asegurar que son verdaderamente distintivas. El examen de los perfiles de los dos conglomerados revela dos conglomerados que son casi uno el espejo del otro. El conglomerado 1 tiene mayores valores sobre X_1 y X_3, mientras que el conglomerado 2 tiene los mayores valores para X_2, X_4, X_6 y X_7. Sólo X_5 no muestra diferencias significativas entre los conglomerados. Si los resultados se restringen sólo a esta solución, los resultados sugerirían dos segmentos bastante diferentes. Los conglomerados no sólo pueden caracterizarse teniendo todos los valores bajos o todos los valores altos, sino que pueden ser evaluados tanto desde un punto de vista conceptual como práctico.

También se muestran en la Tabla 9.7 los perfiles para la solución de cuatro conglomerados. El aumento del número de conglomerados proporciona una estructura bien definida y más variación en términos de las variables de obtención de conglomerados. Examinando los perfiles de los cuatro conglomerados se revela un número de pautas de valores altos frente a bajos. Otro aspecto que varía de la solución de dos conglomerados es que todas las variables de obtención de conglomerados difieren de una manera estadísticamente significativa para los cuatro grupos. Compárese esto con las cinco variables de la solución de dos conglomerados que eran significativamente diferentes en el nivel 0,01. El mayor número de conglomerados muestra una mejora en la representación de los diferentes grupos que puede reflejar una estructura subyacente.

Evaluación de los resultados no jerárquicos Todos los indicadores de esta etapa (regla de parada, ausencia de atípicos y perfiles diferenciados) apoyan las dos soluciones, la de dos y la de cuatro conglomerados. Por tanto, ambas soluciones se pasan al análisis no jerárquico (Paso 2) para obtener las soluciones finales de conglomerado. La selección entre las soluciones de dos o cuatro conglomerados se basarán en las evaluaciones hechas en los pasos cinco y seis.

Paso 2: *Análisis cluster no jerárquico*

El segundo paso utiliza las técnicas no jerárquicas para ajustar los resultados de los procedimientos jerarquizados. Al desarrollar el análisis *cluster*, el investigador deberá tomar los puntos de semilla iniciales de los resultados del Paso 1: en este caso, existen tres centroides de grupo sobre siete variables de obtención de conglomerados (X_1 a X_7) tanto para las soluciones de dos como para las de cuatro conglomerados. Los resultados para ambas soluciones *cluster* (valores del centroide y tamaño del conglomerado) se muestran en la Tabla 9.8. Igual que se observa en los métodos jerárquicos, la solución de dos conglomerados resulta en grupos de casi el mismo tamaño (52 frente a 48 observaciones), y los perfiles se corresponden bastante bien con los perfiles de los conglomerados del procedimiento jerárquico. De nuevo, sólo X_5 no muestra diferencias significativas

TABLA 9.5. Esquema de aglomeración de el análisis *cluster* jerárquico utilizando el método de Ward

	Conglomerado combinado			*Paso de primera aparación del conglomerado*		
Paso	*Conglomerado 1*	*Conglomerado 2*	*Coeficiente de Aglomeración*	*Conglomerado 1*	*Conglomerado 2*	*Siguiente*
1	15	20	0,000	0	0	60
2	5	42	0,005	0	0	94
3	24	27	0,010	0	0	74
4	47	61	0,020	0	0	78
5	19	28	0,040	0	0	60
6	67	90	0,070	0	0	39
7	18	92	0,105	0	0	65
8	51	77	0,140	0	0	72
9	33	62	0,175	0	0	63
10	36	41	0,210	0	0	45
11	85	87	0,260	0	0	69
12	65	79	0,310	0	0	68
13	43	46	0,360	0	0	76
14	25	44	0,410	0	0	63
15	38	63	0,475	0	0	54
16	69	81	0,555	0	0	52
17	94	98	0,650	0	0	73
18	56	91	0,745	0	0	66
19	50	72	0,840	0	0	52
20	75	99	0,950	0	0	62
21	1	95	1,060	0	0	72
22	16	73	1,170	0	0	61
23	37	48	1,280	0	0	58
24	11	100	1,405	0	0	69
25	4	89	1,545	0	0	62
26	84	88	1,685	0	0	45
27	2	83	1,825	0	0	82
28	29	78	1,965	0	0	61
29	3	71	2,105	0	0	75
30	23	32	2,245	0	0	66
31	17	64	2,435	0	0	83
32	12	76	2,650	0	0	67
33	8	68	2,865	0	0	70
34	9	74	3,130	0	0	55
35	52	60	3,420	0	0	57
36	10	34	3,755	0	0	43
37	26	59	4,105	0	0	64
38	49	97	4,525	0	0	81
39	7	67	4,995	0	6	77
40	13	21	5,515	0	0	51
41	82	93	6,040	0	0	91
42	40	54	6,565	0	0	53
43	10	30	7,097	36	0	50
44	66	80	7,632	0	0	59
45	36	84	8,189	10	26	70
46	22	55	8,749	0	0	71
47	6	70	9,409	0	0	57
48	45	86	10,239	0	0	53
49	39	96	11,079	0	0	68

TABLA 9.5. *(Continuación)*

	Conglomerado combinado			*Paso de primera aparación del conglomerado*		
Paso	*Conglomerado 1*	*Conglomerado 2*	*Coeficiente de Aglomeración*	*Conglomerado 1*	*Conglomerado 2*	*Siguiente*
50	10	53	11,965	43	0	56
51	13	35	13,025	40	0	71
52	50	69	14,468	19	16	65
53	40	45	15,970	42	48	73
54	14	38	17,558	0	15	59
55	9	58	19,213	34	0	67
56	10	31	21,261	50	0	58
57	6	52	23,516	47	35	88
58	10	37	25,869	56	23	75
59	14	66	28,244	54	44	80
60	15	19	30,704	1	5	77
61	16	29	33,179	22	28	78
62	4	75	35,714	25	20	74
63	25	33	38,537	14	9	64
64	25	26	41,568	63	37	84
65	18	50	44,879	7	52	76
66	23	56	48,546	30	18	87
67	9	12	52,279	55	32	80
68	39	65	56,214	49	12	89
69	11	85	60,252	24	11	87
70	8	36	64,364	33	45	83
71	13	22	68,580	51	46	90
72	1	51	73,083	21	8	84
73	40	94	77,887	53	17	85
74	4	24	82,785	62	3	82
75	3	10	88,133	29	58	79
76	18	43	93,522	65	13	92
77	7	15	98,977	39	60	86
78	16	47	104,835	61	4	90
79	3	57	111,625	75	0	91
80	9	14	118,530	67	59	81
81	9	49	126,007	80	38	86
82	2	4	134,773	27	74	85
83	8	17	143,875	70	31	88
84	1	25	156,719	72	64	92
85	2	40	170,259	82	73	89
86	7	9	185,590	77	81	94
87	11	23	201,110	69	66	93
88	6	8	218,441	57	83	93
89	2	39	236,111	85	68	96
90	13	16	258,731	71	78	95
91	3	82	281,428	79	41	97
92	1	18	305,027	84	76	95
93	6	11	333,081	88	87	96
94	5	7	364,898	2	86	98
95	1	13	398,082	92	90	98
96	2	6	446,283	89	93	97
97	2	3	522,981	96	91	99
98	1	5	614,954	95	94	99
99	1	2	994,752	98	97	0

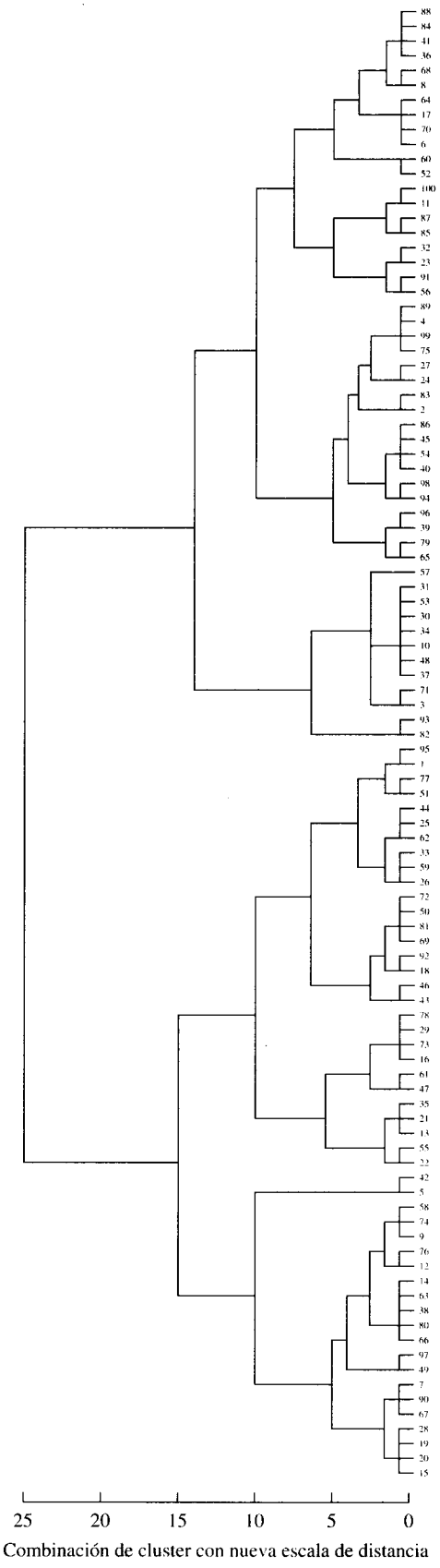


FIGURA 9.11. Dendrograma para el análisis *cluster* jerárquico utilizando el método de Ward.

Tabla 9.7. Perfiles de las variables de obtención de conglomerados para soluciones *cluster* de dos a cuatro variables para el análisis *cluster* jerárquico

PERFILES DE LAS VARIABLES PARA LA ELABORACIÓN DE CONGLOMERADOS

	Valores medios de las variables para obtener conglomerados							
Conglomerado	X_1 *Velocidad de entrega*	X_2 *Nivel de precio*	X_3 *Flexibilidad de precios*	X_4 *Imagen de productor*	X_5 *Servicio conjunto*	X_6 *Imagen fuerza ventas*	X_7 *Calidad producto*	*Tamaño conglomerado*
Solución de dos conglomerados								
1	4,460	1,576	8,900	4,926	2,992	2,510	5,904	50
2	2,570	3,152	6,888	5,570	2,840	2,820	8,038	50
Solución de cuatro conglomerados								
1	4,207	1,624	8,597	4,372	2,879	2,014	5,124	29
2	2,213	2,834	7,166	5,358	2,505	2,689	7,968	38
3	3,700	4,158	6,008	6,242	3,900	3,233	8,258	12
4	4,810	1,510	9,319	5,690	3,148	3,195	6,981	21

CONTRASTE DE SIGNIFICACIÓN DE LAS DIFERENCIAS ENTRE LOS CENTROS DE LOS CONGLOMERADOS

Variables	*Media al cuadrado conglomerado*	*Grados de Libertad*	*Error medio al cuadrado*	*Grados de libertad*	*Valor F*	*Significación*
Solución de dos conglomerados						
X_1 Velocidad de entrega	89,302	1	0,851	98	104,95	0,000
X_2 Nivel de precios	62,094	1	0,811	98	76,61	0,000
X_3 Flexibilidad de precios	101,204	1	0,909	98	111,30	0,000
X_4 Imagen de productos	10,368	1	1,187	98	8,73	0,004
X_5 Servicio conjunto	0,578	1	0,564	98	1,02	0,314
X_6 Imagen de la fuerza de ventas	2,402	1	0,576	98	4,17	0,044
X_7 Calidad del producto	113,849	1	1,377	98	82,68	0,000
Solución de cuatro conglomerados						
X_1 Velocidad de entrega	37,962	3	0,613	96	61,98	0,000
X_2 Nivel de precios	26,082	3	0,659	96	39,56	0,000
X_3 Flexibilidad de precios	39,927	3	0,735	96	54,34	0,000
X_4 Imagen de productos	12,884	3	0,917	96	14,04	0,000
X_5 Servicio conjunto	6,398	3	0,382	96	16,75	0,000
X_6 Imagen de la fuerza de ventas	7,367	3	0,383	96	19,26	0,000
X_7 Calidad del producto	52,203	3	0,960	96	54,37	0,000

entre los conglomerados. Para la solución de cuatro grupos, los tamaños de conglomerado eran bastante similares a aquellas del procedimiento jerárquico, variando en tamaño como mucho en sólo cuatro observaciones. También, los perfiles de conglomerado se ajustan bastante bien. La correspondencia y estabilidad de las dos soluciones de conglomerado entre los métodos jerárquicos y no jerárquicos confirman los resultados sujetos a aceptación práctica y teórica.

Quinto paso: Interpretación de los conglomerados

En la Tabla 9.8 se ofrece información esencial para las fases de interpretación y obtención de perfiles. Para cada conglomerado, se proporciona el centroide de cada uno de las siete variables de obtención de conglomerados junto con los ratios de la F univariante y niveles de significación comparando las diferencias entre las medias de los conglomerados. Mientras en el cuarto paso examinamos los conglomerados en busca de particularidades, ahora considerados la significación práctica de los conglomerados en el cumplimiento de los objetivos de segmentación de los mercados. ¿Cuál es la mejor segmentación del mercado, dos o cuatro segmentos?

Al evaluar los perfiles de las variables de obtención de conglomerados, debemos recordar alguno de los resultados del análisis factorial desarrollados sobre estas siete variables (véase Capí-

TABLA 9.8. Soluciones de dos a cuatro conglomerados del análisis *cluster* no jerárquico con puntos de semilla iniciales desde los resultados jerárquicos

SOLUCIÓN DE DOS CONGLOMERADOS

	Valores medios							
Conglomerado	X_1 *Velocidad de entrega*	X_2 *Nivel de precio*	X_3 *Flexibilidad de precios*	X_4 *Imagen de productor*	X_5 *Servicio conjunto*	X_6 *Imagen fuerza ventas*	X_7 *Calidad producto*	*Tamaño conglomerado*
Centros de conglomerado finales								
1	4,383	1,581	8,900	4,925	2,958	2,525	5,904	52
2	2,575	3,212	6,804	5,598	2,871	2,817	8,127	48
Significación estadística de las diferencias de conglomerados								
Valor *F*	87,72	86,75	133,18	9,60	0,33	3,67	96,40	
Significación	0,000	0,000	0,000	0,003	0,566	0,058	0,000	

	Obtención de perfiles de los conglomerados			
	Conglomerado			
Validación predictiva	1	2	*Valor F*	*Significación*
X_9 Nivel de uso	49,212	42,729	14,789	0,000
X_{10} Satisfacción	5,133	4,379	23,826	0,000

SOLUCIÓN PARA CUATRO CONGLOMERADOS

	Valores medios							
Conglomerado	X_1 *Velocidad de entrega*	X_2 *Nivel de precio*	X_3 *Flexibilidad de precios*	X_4 *Imagen de productor*	X_5 *Servicio conjunto*	X_6 *Imagen fuerza ventas*	X_7 *Calidad producto*	*Tamaño conglomerado*
Centros de conglomerado finales								
1	4,094	1,621	8,630	4,415	2,830	2,079	5,273	33
2	2,171	2,846	7,123	5,403	2,489	2,683	8,194	35
3	3,662	4,200	5,946	6,123	3,900	3,177	7,946	13
4	4,884	1,511	9,368	5,811	3,179	3,300	7,000	19
Significación estadística de las diferencias de conglomerados								
Valor *F*	56,35	46,71	67,86	14,60	18,60	19,71	57,60	
Significación	0,000	0,000	0,000	0,000	0,000	0,000	0,000	

	Obtención de perfiles de los conglomerados					
	Conglomerado					
Validación predictiva	1	2	3	4	*Valor F*	*Significación*
X_9 Nivel de uso	46,333	41,229	46,769	54,211	11,304	0,000
X_{10} Satisfacción	4,839	4,134	5,038	5,642	22,212	0,000

tulo 3). Al definir los factores que son «dimensiones» de las variables que están altamente intercorrelacionadas, surgen dos factores. El primer factor contiene X_1, X_2, X_3 y X_7, y el segundo factor contiene los dos elementos de imagen, X_4 y X_6. En el primer factor, X_2 y X_7 están inversamente relacionados con X_1 y X_3, lo que significa que cuando uno aumenta, el otro debería disminuir. A nuestros efectos, esto sugiere que altas puntuaciones en X_1 y X_3 deberían estar asociados con bajas puntuaciones de X_2 y X_7. Por tanto, esperar que en un conglomerado sean todas altas o bajas, de tal forma que cuando hablamos de que un conglomerado sea «alto» sobre este conjunto de variables, significa altas puntuaciones para X_1 y X_3 y valores bajos para X_2 y X_7. Otra posibilidad es calcular puntuaciones compuestas para cada dimensión (véase Capítulo 3 para más detalles), pero en nuestro ejemplo sólo calculamos perfiles para variables individuales.

Solución de dos conglomerados

Comenzamos examinando los niveles de significación para las diferencias entre los conglomerados. Nótese que seis de los siete conglomerados son estadísticamente significativos (véase Tabla 9.8). Sólo X_5, servicio conjunto, no es significativamente diferente entre los dos conglomerados. Al formar un perfil de los conglomerados (véase Figura 9.12), vemos que el conglomerado 1 es significativamente mayor en esa primera «dimensión» de variables, mientras que el conglomerado 2 tiene mayores percepciones de HATCO sobre las dos variables imagen (X_4 y X_6). Uno debería notar sin embargo, que las diferencias son más pronunciadas en el primer conjunto de variables y que las variables de imagen tiene substancialmente menos delineación entre los conglomerados, incluso aunque sean estadísticamente significativas. Con esto se debería centrar la atención práctica sobre la cuarta variable en el primer conjunto y relegar las variables de imagen a un papel secundario, aunque sean las variables descriptores claves del segundo conglomerado.

Solución de cuatro conglomerados

La solución de cuatro conglomerados representa en realidad la partición de cada uno de los conglomerados de la solución de dos conglomerados. El conglomerado 1 se divide en los conglomerados 1 y 4, mientras que el conglomerado 2 se divide en los conglomerados 2 y 3. Al formular una interpretación de los perfiles, surgen varias pautas. Dado que los conglomerados 1 y 4 emergen del conglomerado 1 de la solución de dos conglomerados, debería compartir al menos algunas pautas similares. Así ocurre en el primer conjunto de variables (X_1, X_2, X_3 y X_7), manteniendo su diferen-

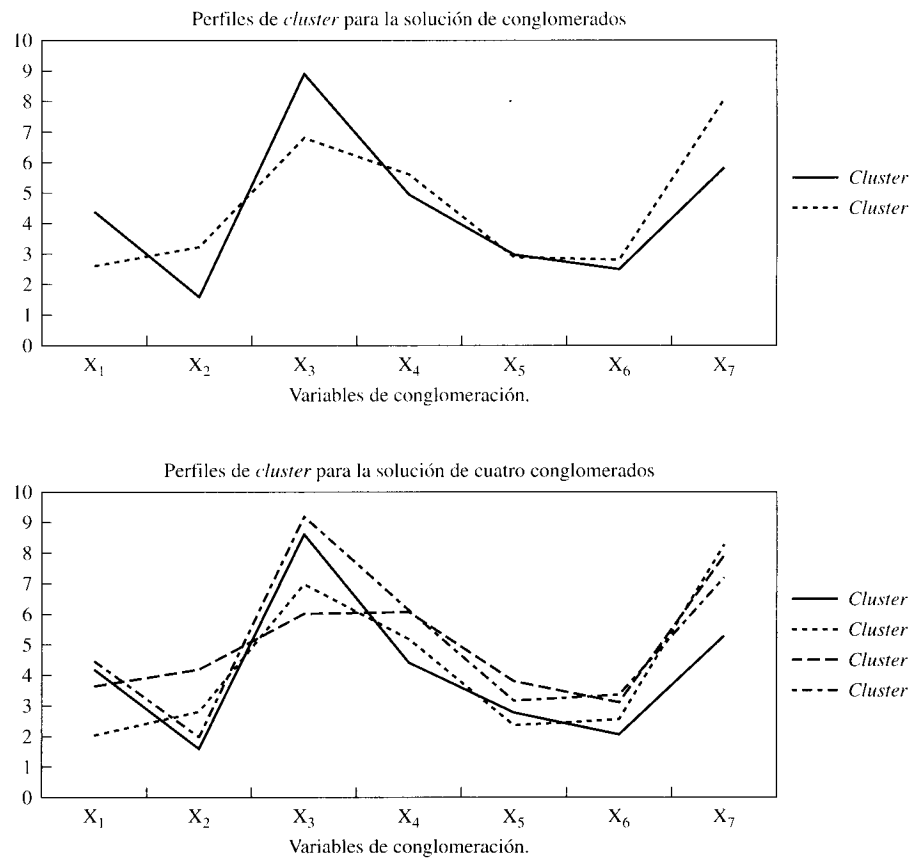


FIGURA 9.12. Perfiles gráficos de la solución de conglomerados de dos y cuatro variables del análisis *cluster* no jerárquico.

cia frente a los conglomerados 2 y 3. Sus perfiles varían ligeramente, donde el conglomerado 1 es mayor en X_2 y el conglomerado 4 en X_1, X_3 y X_7. Las marcadas diferencias vienen de las variables imagen, donde el conglomerado 1 sigue la pauta vista en la solución de dos conglomerados con bajas puntuaciones mientras que el conglomerado 4 contrarresta esta tendencia al tener puntuaciones de imagen altamente positivas. Con ella se representa la diferencia más marcada entre los conglomerados 1 y 4. El conglomerado 4 tiene percepciones relativamente altas de HATCO para todas las variables, mientras que el conglomerado 1 es mayor sólo en el primer conjunto de variables.

Los conglomerados restantes, 2 y 3, siguen la pauta de tener puntuaciones relativamente altas sobre las variables imagen y bajas en el resto. Pero también se diferencian, en la medida en que existe separación entre ellos sobre el primer conjunto de variables, particularmente X_1, X_2 y X_3. La identificación de estas pautas entre las variables de obtención de conglomerados permite al investigador desarrollar perfiles más profundos y desarrollar tácticas dirigidas hacia las percepciones específicas de cada grupo.

Sexto paso: Validación y perfil de los grupos

En este paso final, los procesos de validación y formación de perfiles son críticos debido al carácter exploratorio y a veces *ateórico* del análisis *cluster*. Es esencial que el investigador lleve a cabo todos los tests para confirmar la validez de la solución *cluster* a la vez que se asegura que la solución *cluster* tiene significación práctica. Los investigadores que minimizan o se saltan este paso se exponen al riesgo de aceptar una solución que es específica sólo para la muestra y tiene una generalización limitada o incluso reducida cuando se utiliza más allá de la mera descripción de los datos sobre las variables de obtención de conglomerados.

Validación de las soluciones cluster

El proceso de validación se realiza en dos pasos. En primer lugar, la validez se evalúa mediante la aplicación de métodos *cluster* alternativos y comparando las soluciones. A continuación se evalúan los conglomerados a efectos de la validez predictiva sobre dos medidas adicionales (X_9 nivel de utilización y X_{10} nivel de satisfacción) que son indicativas del potencial de estrategias de diferenciación entre los conglomerados (véase Tabla 9.8).

Aplicación de un segundo análisis jerárquico Como primera prueba de la validez de la estabilidad de la solución *cluster*, se lleva a cabo un segundo análisis no jerárquico, esta vez permitiendo que el proceso seleccione aleatoriamente los puntos de semilla iniciales para ambas soluciones *cluster*. Los resultados de la Tabla 9.9 confirman una consistencia de los resultados para ambas soluciones. Los tamaños de los conglomerados son comparables para cada solución, y los perfiles de los conglomerados son muy similares. Otra aproximación, aunque no se considera aquí, es emplear medidas de similitud diferentes y métodos de encadenamiento en el análisis jerárquico para evaluar la consistencia de los resultados. Pero dada la estabilidad de los resultados entre los puntos de semilla especificados y la selección aleatoria, la dirección de HATCO debería sentirse confiada en que existen «verdaderas» diferencias entre los clientes en términos de sus pedidos a un proveedor y que la estructura diseñada en el análisis *cluster* está empíricamente contrastada.

Evaluación de la validez predictiva Para evaluar la validez predictiva, nos centramos en variables que tiene una relación basada teóricamente para las siete variables de conglomerado, pero que no fueron incluidas en la solución *cluster*. En este ejemplo, consideramos las variables X_9 (nivel de uso) y X_{10} (nivel de satisfacción) como se muestra en la Tabla 9.8. Se ha mostrado que estas dos variables tienen una relación definida con las variables de elaboración de conglomerados en anteriores técnicas multivariantes, en particular la correlación canónica y regresión múltiple. Dada esta relación, deberíamos ver diferencias significativas en estas variables a lo largo de los con-

TABLA 9.9. Perfiles de variables de obtención de conglomerados para las soluciones de dos y cuatro conglomerados del análisis *cluster* no jerárquico con puntos de semilla aleatorios

PERFILES DE VARIABLES DE ELABORACIÓN DE CONGLOMERADOS

Valores medios de las variables de elaboración de conglomerados

Conglomerado	X_1 *Velocidad de entrega*	X_2 *Nivel de precio*	X_3 *Flexibilidad de precios*	X_4 *Imagen de productor*	X_5 *Servicio conjunto*	X_6 *Imagen fuerza ventas*	X_7 *Calidad producto*	*Tamaño conglomerado*
Solución de dos conglomerados								
Puntos de semilla iniciales								
1	4,100	0,600	6,900	4,700	2,400	2,300	5,200	
2	1,800	3,000	6,300	6,600	2,500	4,000	8,400	
Solución de *cluster* final								
1	4,383	1,581	8,900	4,925	2,958	2,525	5,904	52
2	2,575	3,212	6,804	5,598	2,871	2,817	8,127	48
Solución de cuatro conglomerados								
Puntos de semilla iniciales								
1	4,100	0,600	6,900	4,700	2,400	2,300	5,200	
2	1,800	3,000	6,300	6,600	2,500	4,000	8,400	
3	3,400	5,200	5,700	6,000	4,300	2,700	8,200	
4	2,700	1,000	7,100	5,900	1,800	2,300	7,800	
Solución de *cluster* final								
1	4,144	1,579	8,635	4,418	2,835	2,088	5,315	34
2	2,024	2,766	6,941	5,162	2,366	2,555	8,269	29
3	3,614	4,129	5,950	6,064	3,843	3,164	7,950	14
4	4,404	1,943	9,183	6,087	3,165	3,352	7,187	23

glomerados. Si existen variables significativas sobre estas dos variables, podemos deducir la conclusión de que los conglomerados diseñan grupos que tienen validez predictiva.

Para la solución de dos conglomerados, los ratios univariantes F muestran que las medias de los conglomerados para ambas variables son significativamente diferentes. Los procesos de obtención de perfiles muestran aquí que los clientes del conglomerado 1, que califican a HATCO más alto para el primer conjunto de variables, tenían mayores niveles de uso y satisfacción con HATCO, como sería de esperar. De la misma forma, los clientes del conglomerado 2 tenían calificaciones más bajas para el primer conjunto de variables y calificaciones más bajas para las dos variables adicionales. Aunque no es el propósito del análisis *cluster* identificar la relación entre las variables de obtención de conglomerado y las variables adicionales, los resultados aquí apoyan las relaciones halladas con otras técnicas multivariantes.

La solución de cuatro conglomerados muestra una pauta similar, donde los conglomerados muestran diferencias estadísticamente significativas sobre estas dos variables adicionales. Aunque está más lejos del alcance de este análisis relacionar las variables de obtención de conglomerados para estos niveles de uso y satisfacción, las pautas de las diferencias entre los conglomerados ofrecen suficiente evidencia como para deducir la conclusión de que la solución de cuatro conglomerados también tiene un nivel adecuado de validez predictiva.

Obtención de perfiles de las soluciones cluster *sobre variables adicionales*

La tarea final es obtener perfiles de los conglomerados sobre un conjunto de variables adicionales no incluidas en el procedimiento de obtención de conglomerados o utilizarlo para evaluar la validez predictiva. En este ejemplo, se ofrecen cierto número de características de los clientes de HATCO. Estas incluyen X_8 (tamaño de la empresa), X_{11} (términos de compra), X_{12} (estructura de adquisición), X_{13} (tipo de industria) y X_{14} (situación de compra). La Tabla 9.10 ofrece un perfil descriptivo de ambas soluciones *cluster* sobre estas características. Como podemos ver, las solucio-

TABLA 9.10 Perfil de las soluciones de dos y cuatro conglomerados sobre características adicionales (X_8, X_{11}, X_{12}, X_{13} y X_{14})

	Perfiles de los conglomerados			
	Conglomerado		*Chi-cuadrado de Pearson*	*Significación*
X_8 *Tamaño de la empresa*	1	2		
Pequeña	50	10	59,001	0,000
Grande	2	38		
X_{11} *Términos de compra*				
Términos de compra	2	38	59,001	0,000
Análisis de valor total	50	10		
X_{12} *Estructura de adquisición*				
Descentralizada	50	0	92,308	0,000
Centralizada	2	48		
X_{13} *Tipo de industria*				
Categoría SIC uno	28	22	0,641	0,423
Categoría SIC dos	24	26		
X_{14} *Tipo de situación de compra*				
Nueva compra	10	24	36,634	0,000
Recompra modificada	10	22		
Recompra directa	32	2		

	Conglomerado				*Chi-cuadrado de Pearson*	*Significación*
X_8 *Tamaño de la empresa*	1	2	3	4		
Pequeña	31	6	4	19	59,919	0,000
Grande	2	29	9	0		
X_{11} *Términos de compra*						
Términos de compra	2	29	9	0	59,919	0,000
Análisis de valor total	31	6	4	19		
X_{12} *Estructura de adquisición*						
Dencentralizada	31	0	0	19	92,485	0,000
Centralizada	2	35	13	0		
X_{13} *Tipo de industria*						
Categoría SIC uno	23	17	5	5	10,105	0,018
Categoría SIC dos	10	18	8	14		
X_{14} *Tipo de situación de compra*						
Nueva compra	8	23	1	2	62,189	0,000
Recompra modificada	8	10	12	2		
Recompra directa	17	2	0	15		

nes de dos y cuatro conglomerados tienen perfiles diferenciados sobre este conjunto de variables adicionales. Sólo X_{13} no muestra diferencias para la solución de dos conglomerados, teniendo todas las variables diferencias significativas para las soluciones de dos y cuatro conglomerados. Por ejemplo, el conglomerado 1 de la solución de dos conglomerados puede caracterizarse por pequeñas empresas embarcadas en repetidas compras a HATCO que fundamentalmente utilizan el análisis del valor total en un sistema de adquisición descentralizada.

La importancia de identificar perfiles únicos en estos conjuntos de variables adicionales está en la evaluación tanto de su significación práctica como teórica. Al evaluar la significación práctica, el investigador puede exigir que los conglomerados muestren diferencias sobre un conjunto de variables adicionales. En nuestro ejemplo, un buen análisis de segmentación no sólo requiere la identificación de grupos homogéneos (conglomerados), sino que también sean identificables (únicamente descritos sobre otras variables). Cuando se utiliza el análisis *cluster* para verificar una tipología u otra agrupación propuesta de objetos, variables asociadas, sean antecedentes o resultados, pueden perfilarse para asegurar la correspondencia de las agrupaciones dentro de un gran modelo teórico.

Una visión práctica

El análisis *cluster* de 100 clientes de HATCO tuvo éxito en la realización de una segmentación de mercado de los clientes de HATCO. No sólo creó agrupaciones homogéneas de clientes basándose en sus percepciones de HATCO, sino que también encontró que estos conglomerados cumplen los test de validez predictiva y diferenciación sobre un conjunto de variables adicionales, todas necesarias para obtener la significación práctica. Los segmentos representan perspectivas muy diferentes de los clientes de HATCO, variando tanto en el tipo de variables vistas más positivamente como la magnitud de las percepciones.

Un tema no resuelto hasta este punto es la selección de la solución cluster definitiva entre los dos y los cuatro conglomerados. Ambas soluciones cluster ofrecen segmentos de mercado viables, en la medida en que son de un tamaño substancial y no contienen tamaños de conglomerados reducidos debido a los atípicos. Además, cumplen todos los criterios para buena segmentación de mercados. En cada caso, los conglomerados (segmentos de mercado) representan conjuntos de consumidores con percepciones homogéneas que pueden ser identificadas singularmente, siendo por tanto candidatos óptimos a programas de márketing diferenciados. El investigador puede utilizar la solución de dos conglomerados para ofrecer una delimitación básica de los clientes que varían en comportamiento y percepciones de compra, o emplear la solución de cuatro conglomerados para una estrategia de segmentación más compleja que ofrece una combinación altamente diferenciada de percepciones de clientes, así como opciones para fijar objetivos.

Resumen

El análisis *cluster* puede ser muy útil como técnica de reducción de datos. Pero dado que su aplicación es más un arte que una ciencia, se puede abusar fácilmente o aplicar mal por parte de los investigadores. Diferentes algoritmos y medidas entre objetos pueden afectar a los resultados. La selección del conglomerado de la solución cluster final se basa en la mayoría de los casos tanto en consideraciones objetivas como subjetivas. El investigador prudente, por tanto, considera estos temas y siempre evalúa el impacto de todas las decisiones.

El análisis *cluster*, junto con el análisis multidimensional, debido a su falta de base estadística para inferir de la población, tiene una mayor necesidad de aplicarse varias veces bajo condiciones cambiantes. Si el investigador procede con cautela, sin embargo, el análisis cluster puede ser un instrumento valioso en la identificación de pautas latentes mediante la sugerencia de agrupaciones (conglomerados) de objetos que no son discernibles mediante otras técnicas multivariantes.

Cuestiones de revisión

1. ¿Cuáles son los pasos básicos en la aplicación de análisis cluster?
2. ¿Cuál es el propósito del análisis cluster y cuándo debería utilizarse en lugar del análisis factorial?
3. ¿Qué debería considerar el investigador cuando selecciona una medida de similitud para utilizarla en el análisis cluster?
4. ¿Cómo sabe el investigador si tiene que utilizar técnicas cluster jerárquicas o no jerárquicas? ¿Bajo qué condiciones utilizaría cada aproximación?
5. ¿Cómo decide el investigador el número de conglomerados que tiene su solución?
6. ¿Cuál es la diferencia entre el paso de interpretación y el paso de obtención de perfiles?
7. ¿Cómo utilizan los investigadores las representaciones gráficas del procedimiento cluster?

Bibliografía

1. Aldenderfer, Mark S., and Roger K. Blashfield (1984), *Cluster Analysis.* Thousand Oaks, Calif.: Sage Publications.
2. Anderberg, M. (1973), *Cluster Analysis for Applications.* New York: Academic Press.
3. Bailey, Kenneth D. (1994), *Typologies and Taxonomies: An Introduction to Classification Techniques.* Thousand Oaks, Calif.: Sage Publications.
4. Everitt, B. (1980), *Cluster Analysis*, 2d ed. New York: Halsted Press.
5. Green, P. E. (1978), *Analyzing Multivariate Data.* Hinsdale, Ill.: Holt, Rinehart & Winston.
6. Green, P. E., and J. Douglas Carroll (1978), *Mathematical Tools for Applied Multivariate Analysis.* New York: Academic Press.
7. McIntyre, R. M., and R. K. Blashfield (1980), «A Nearest-Centroid Technique for Evaluating the Minimum-Variance Clustering Procedure.» *Multivariate Behavioral Research* 15: 225-38.
8. Milligan, G. (1980), «An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms.» *Psychometrika* 45 (September): 325-42.
9. Milligan, Glenn W., and Martha C. Cooper (1985), «An Examination of Procedures for Determining the Number of Clusters in a Data Set.» *Psychometrika* 50(2): 159-79.
10. Morrison, D. (1967), «Measurement Problems in Cluster Analysis.» *Management Science* 13 (August): 775-80.
11. Overall, J. (1964), «Note on Multivariate Methods for Profile Analysis.» *Psychological Bulletin* 61(3): 195-98.
12. Punj, G., and D. Stewart (1983), «Cluster Analysis in Marketing Research: Review and Suggestions for Application.» *Journal of Marketing Research* 20 (May): 134-48.
13. Rohlf, F. J. (1970), «Adaptive Hierarchical Clustering Schemes.» *Systematic Zoology* 19: 58.
14. Schaninger, C. M., and W. C. Bass (1986), «Removing Response-Style Effects in Attribute-Determinance Ratings to Identify Market Segments.» *Journal of Business Research* 14: 237-52.
15. Shephard, R. (1966), «Metric Structures in Ordinal Data.» *Journal of Mathematical Psychology* 3: 287-315.
16. Sneath, P. H. A., and R. R. Sokal (1973), *Numerical Taxonomy.* San Francisco: Freeman Press.

LECTURA COMPLEMENTARIA

*Aplicación del análisis cluster**

AZUCENA PENELAS
Universidad de Alcalá de Henares

CONSIDERACIONES PREVIAS

Una técnica para la formación de grupos buscando la optimización es el análisis cluster. En el ejemplo que vamos a exponer, de las distintas aproximaciones posibles para obtener grupos mediante esta técnica, se ha elegido un método de agrupamiento jerárquico. Inicialmente se empieza con un número de «clusters» compuesto por un solo sujeto; busca el par más cercano de «clusters» y los une formando un nuevo y único grupo, continuando de esta forma hasta que se forma un solo cluster que contiene a todos los sujetos originales. La búsqueda del par más cercano requiere comparar entre sí todas las distancias euclídeas, por lo que este algoritmo sólo es aplicable en casos moderadamente grandes.[1]

El paquete estadístico empleado en el análisis es el SYSTAT.[2] Este programa ofrece dos clases distintas de cluster, definidos por los comandos JOIN (agrupamiento jerárquico) y KMEANS (método de partición no necesariamente jerárquico). Se ha utilizado el JOIN, que opera sobre un fichero rectangular o sobre uno que contenga una matriz simétrica.

Una vez que se ha elegido el algoritmo, se debe decidir sobre la medida de similitud entre los objetos. Entre las distintas alternativas, se ha optado por la distancia euclídea (raíz cuadrada de las distancias medias al cuadrado), ya que las variables se encuentran medidas en una escala común, sin ser necesario que estas se estandaricen previamente.

El siguiente paso ha consistido en la elección del método de encadenamiento. El método escogido es el de encadenamiento completo o el «vecino más lejano», que calcula la distancia entre dos cluster como la distancia entre sus dos puntos más distantes. El encadenamiento completo tiende a producir cluster más compactos[3].

El SYSTAT ofrece distintas posibilidades de exponer los resultados obtenidos por el algoritmo JOIN. Para la unión de filas o columnas, imprime un árbol (o dendograma), que facilita la visualización de los grupos.

El árbol se imprime con un único orden en el cual cada rama es alineada para que los objetos más similares estén más cercanos. El algoritmo para ordenar los árboles es el de Gruvaeus y Wainer[4]. Este orden puede diferir de otros árboles impresos por otros programas de cluster, si éstos no utilizan un algoritmo de serie para determinar cómo ordenar las ramas. Al unir filas, cada una de la ramas finales puede ser etiquetadas con la denominación de un sujeto (caso).

Un problema importante, que surge en el análisis cluster, es decidir cuántos cluster son necesarios para describir adecuadamente tanto la diversidad como la similitud en una población[5]. Algunas de las aplicaciones que sugieren los libros de texto para buscar puntos de inflexión, no son útiles para identificar la mejor solución[6]. Por esta razón, otros autores opinan que debe elegirse según la claridad de las descripciones del cluster y su aplicación práctica. Esta ha sido la opción seguida por esta investigación.

Siguiendo el proceso descrito anteriormente, se ha aplicado el análisis cluster a los datos sobre segmentación de clientes obtenidos sobre una muestra de 52 directivos de marketing de otras tantas entidades financieras (Bancos y Cajas de Ahorro) que en la fecha de su realización operaban en España[7].

[1] Hartigan J. A. *Clustering Algorithms.*, Ed. John Wiley & Sons, 1975, pág. 12.

[2] *SYSTAT: The System for Statistics.*, Evanstons, IL., ed. SYSTAT, Inc. 1990.

[3] SYSTAT *ob.cit.*, pág. 31.

[4] Gruvaeus, G. and H. Wainer. «Two aditions to hierarchical cluster analysis». The British Journal of mathematical and Statistical Psychology. 25. 1972, págs. 200-206.

[5] Saunders, J. «Cluster analysis for market segmentation». *European Journal of marketing.* 14, 1980, págs. 422-435.

[6] Hooley, G.J., J.E. Iynch and D. Jobber. «Generic marketing strategies» International Journal of Research in Marketing, 9(1), 1992, págs. 75-89.

[7] Este apartado forma parte de un estudio más amplio realizado en la elaboración de la Tesis Doctoral titulada «Estrategia Comercial en el Sistema Bancario Español: Factores determinantes y tipología». Aut. Azucena Penelas Leguía. Universidad de Alcalá, 1994.

* Esta aplicación ha sido desarrollada específicamente para este libro. Por tanto, manifestamos nuestro especial agradecimiento a su autora (N. de la R.T.).

TABLA 9.L.1. Grupos por segmentos de clientes

Clientes. Segmentos	*Grupo 1*		*Grupo 2*		*Grupo 3*		*Grupo 4*	
N.º de casos	*2*		*33*		*8*		*9*	
Variables	*Media*	*Vza.*	*Media*	*Vza.*	*Media*	*Vza.*	*Media*	*Vza.*
1. < 2 millones de ptas			3,82	1,03	3,29	2,90	—	—
2. ⩾ 2 a 7millones de ptas			4,22	0,50	3,88	098	—	—
3. > 7 mill. y patrimonio < a 50			4,09	0,34	3,88	0,13	—	—
4. > 7 mill. y patrimonio ⩾ a 50			3,82	0,47	3,75	0,79	—	—
5. Estudiantes	2,00		3,21	1,11	2,50	1,43	—	—
6. Asalariados	4,00	0,00	4,42	0,50	3,25	2,21	—	—
7. Autónomos	4,00	0,00	4,12	0,48	2,88	0,98	—	—
8. Profesionales liberales	4,00	0,00	4,06	0,43	3,63	0,84	—	—
9. Empresarios	4,00	2,00	4,00	0,56	4,13	0,70	—	—
10. Jubilados/pensionistas	3,50	0,50	4,21	0,86	2,75	1,64	—	—
11. Jóvenes solteros	2,50	0,50	3,09	0,71	1,67	1,33	—	—
12. Matrimonio joven sin hijos	4,00	0,00	3,88	0,55	2,33	2,33	—	—
13. Matrimonio con hijos < de 6 años	3,50	0,50	3,97	0,41	1,50	0,50	—	—
14. Matrimonio con hijos ⩾ de 6 a ños	3,50	0,50	4,06	0,37	1,50	0,50	—	—
15. Matrimonio sin hijos	4,00	0,00	4,03	0,34	1,50	0,50	—	—
16. Mayores de 65 años	3,50	0,50	3,82	1,15.	2,80	0,70	—	—
17. Hábitats < 50.000 hab	3,50	0,50	3,79	0,80	2,75	2,92	—	—
18. Hábitats de 50.000 a 200.000 h.	3,50	0,50	4,15	0,38	3,80	0,70	—	—
19. Hábitats > 200.000 a 1 millón	3,50	0,50	4,18	0,59	3,83	0,17	—	—
20. Hábitats > 1 millón de hab.	3,00		3,61	1,37	3,80	0,70	—	—

Escala de valoración: 1= nula, 2 = poca, 3 = media, 4 = mucha, 5 = máxima.

En este apartado se solicita a los directivos una valoración sobre la importancia asignada en la estrategia comercial de su banco a una relación de 20 segmentos predefinidos (variables). La escala de valoración utilizada es de Likert, con cinco categorías (del 1 al 5), que se corresponden con una importancia nula, poca, media, mucha y máxima respectivamente.

RESULTADOS

Los resultados obtenidos tras la aplicación del análisis cluster aparecen en el dendrograma (Figura 9.L.1), y las desviaciones medias de los grupos se recogen en la Tabla 9.L.1.

Para confirmar la claridad de los cluster obtenidos, se ha aplicado un análisis discriminante. Las estimaciones del modelo discriminante aplicado, coincide en el 100 por cien de los casos en cada uno de los cuatro grupos que se han obtenido.

Un primer examen del dendrograma por segmentos (Figura 9.L.1) nos indica que la diferencia entre los grupos es amplia, ya que la unión de los mismos se va realizando a niveles de la escala muy altos, siendo la unión final en el último valor posible, es decir, en el 5.

Los grupos tienen un número de elementos muy dispar, que varía entre los dos del grupo 1, treinta y tres el grupo 2, ocho el grupo 3 y nueve el grupo 4.

El grupo 1 está compuesto por dos cajas de ahorro de tamaño pequeño. En este caso, la ausencia de valoración en los segmentos por nivel de renta (variables 1 a 4), puede indicar que este grupo ignora los segmentos. Por actividad profesional (variables 5 a 10), las valoraciones indican que las entidades de este grupo están interesadas en todas las que se relacionan, aunque menos por los estudiantes. Por ciclo de vida familiar (variables 11 a 16) parecen inclinarse por los matrimonios sin hijos, ya sean o no jóvenes, aunque sin rehusar a los que sí los tienen. Por último en la segmentación por hábitat (variables 17 a 20) no muestran ninguna prioridad, aunque hay una ligera inclinación por aquellos inferiores a un millón de habitantes.

El grupo 2 incluye al mayor número de entidades, en proporción similar de bancos (16) y cajas (17). La mayoría de los grandes bancos y cajas se encuentran incluidos en este grupo aunque junto a ellos existen otras de tamaño inferior, tanto de implantación nacional como regional. A pesar del elevado número (33 entidades), no existe una dispersión elevada en sus valoraciones. El grupo queda formado a una distancia relativamente corta (Figura 9.L.2).

La puntuación recogida en general en cuanto al interés de este grupo por todos los segmentos es alta, lo que parece indicar una estrategia global. La valoración más alta según nivel de renta está en aquellos comprendidos entre 2 y 7 millones de pesetas, pero también es fuerte la importancia asignada al resto. Los asalariados son el grupo profesional más valorado seguidos de cerca por los autónomos, profesionales liberales y empresarios, en todos los casos con valoraciones importantes. No hay tampoco diferencias sustanciales en la importancia, siempre elevada, de los segmentos según el ciclo de vida familiar. En los hábitats, pa-

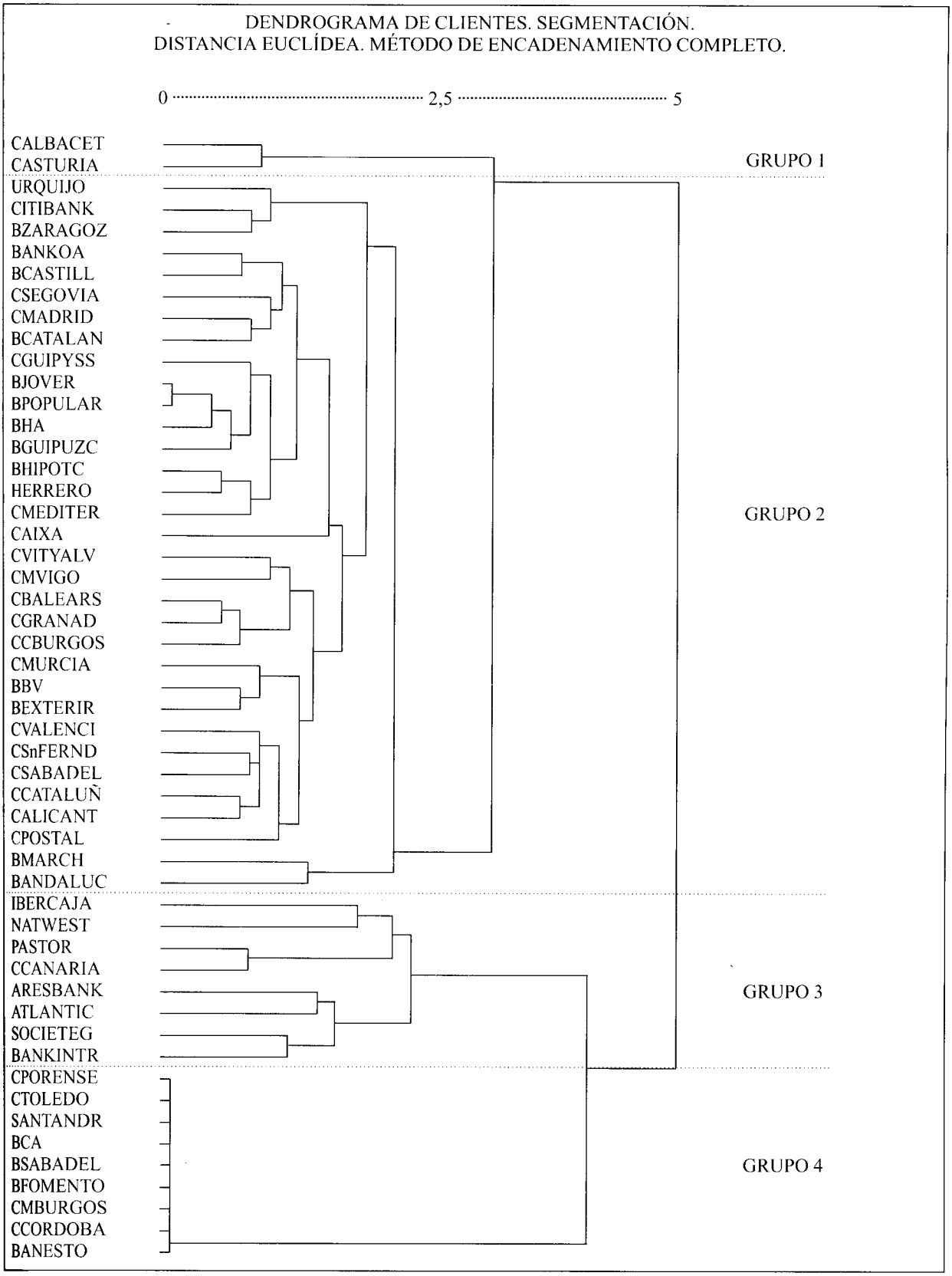

DENDROGRAMA DE CLIENTES. SEGMENTACIÓN.
DISTANCIA EUCLÍDEA. MÉTODO DE ENCADENAMIENTO COMPLETO.
0
2,5
5
CALBACET
CASTURIA
URQUIJO
CITIBANK
BZARAGOZ
BANKOA
BCASTILL
CSEGOVIA
CMADRID
BCATALAN
CGUIPYSS
BJOVER
BPOPULAR
BHA
BGUIPUZC
BHIPOTC
HERRERO
CMEDITER
CAIXA
CVITYALV
CMVIGO
CBALEARS
CGRANAD
CCBURGOS
CMURCIA
BBV
BEXTERIR
CVALENCI
CSnFERND
CSABADEL
CCATALUÑ
CALICANT
CPOSTAL
BMARCH
BANDALUC
IBERCAJA
NATWEST
PASTOR
CCANARIA
ARESBANK
ATLANTIC
SOCIETEG
BANKINTR
CPORENSE
CTOLEDO
SANTANDR
BCA
BSABADEL
BFOMENTO
CMBURGOS
CCORDOBA
BANESTO
GRUPO 1
GRUPO 2
GRUPO 3
GRUPO 4

FIGURA 9.L.1.

rece reflejarse un interés algo mayor por los intermedios entre 50.000 y 1 millón de habitantes. Comparando con el resto de los grupos definidos por el cluster, los segmentos de estudiantes y jóvenes solteros, son más valorados en éste.

El grupo 3 está formado por una mayoría de bancos (6) frente a cajas (2). Los bancos son de tamaño intermedio a pequeño, con un elevado número de bancos extranjeros. Las cajas son dispares en volumen de ATM[8].

Este grupo se caracteriza por su preferencia sobre unos segmentos. Se valoran más los patrimonios entre intermedios y altos, las actividades profesionales favoritas son las de los profesionales liberales y sobre todo empresarios, y en cuanto a los hábitats, se inclinan por los de más de 50.000 habitantes incluyendo los de más de un millón. No consideran importantes los segmentos siguiendo el ciclo de vida familiar ni el resto de segmentos analizados.

Hay que destacar que en este grupo existe una fuerte dispersión respecto a algunos segmentos como el de renta inferior a 2 millones, asalariados y hábitats de menos de 50.000 habitantes, lo que indica que alguna/s de las entidades, ofrece una valoración sensiblemente superior o inferior a la media en estos segmentos.

El grupo 4 queda definido por su falta de valoración a los segmentos. Esto incluye dos tipos de casos diferentes. Uno, el mayoritario, que contesta que no segmenta su mercado de clientes y por tanto no asignan ningún valor a los segmentos. El otro grupo está formado por entidades, que aunque es constatable en la realidad que sí utilizan la segmentación en la elaboración de su estrategia comercial, no quieren indicar sus preferencias, por lo cual no contestan a este apartado de preguntas.

En este grupo se encuentra un número similar de cajas que de bancos, y aunque el volumen de ATM es generalmente bajo, existe alguna entidad de gran tamaño.

El estudio de las variables que más influyen en la diferencia entre grupos ofrece algunos datos interesantes.

Por actividad profesional, los asalariados, autónomos y jubilados tienen especial importancia para los grupos 1 y 2, y sólo media o baja para el resto. Por ciclo de vida familiar, las diferencias vienen marcadas en todos los segmentos por el grupo 3, ya que en ningún caso los considera importantes, al contrario que en los grupos 1 y 2 donde la valoración es siempre superior.

Resumiendo, el grupo 3 muestra mayores discrepancias respecto a los otros dos, a consecuencia de su preferencia hacia sólo algunos segmentos. Los grupos 1 y 2 tienen valores más parecidos y con la misma tendencia en todos los segmentos, a excepción del segmento de estudiantes, que es poco importante para el grupo 1 (además, en este grupo no se valoran los segmentos definidos por renta). Finalmente, el grupo 4 se caracteriza por una ausencia de valoraciones, consecuencia en la mayoría de los casos, de la inexistencia de una política de segmentación.

En base a estos resultados, se puede deducir la existencia de tres estrategias diferentes de aplicación de la segmentación de mercados en la estrategia comercial. Estas estrategias se denominan estrategia diferenciada, concentrada e indiferenciada (Santesmases, 1996, pág. 231).

La estrategia diferenciada es practicada por las entidades financieras del grupo 2 y en gran medida por las del grupo 1, que muestran tener interés por todos los segmentos. La estrategia concentrada se corresponde con las valoraciones de las entidades del grupo 3, al mostrar su interés solo por algunos de los segmentos citados, en los cuales se concentran. La ausencia de valoraciones indica ausencia de segmentación, o lo que es lo mismo una estrategia indiferenciada. Esta es la estrategia seguida por las entidades del grupo 4 (teniendo en cuenta las salvedades comentadas con anterioridad para este grupo).

BIBLIOGRAFÍA

GRUVAEUS, G., y WAINER H.: «Two aditions to hierarchical cluster analysis». *The British Journal of Mathematical and Statistical Psychology* 25. 1972, págs 200-206.

HARTIGAN, J.A.: *Clustering Algorithms.* John Wiley & Sons. 1975, pág.12

HOOLEY, G.J.; IYNCH, J.E., y JOOBER, D.: «Generic marketing strategies». *International Journal of Research in marketing* 9, 1992 págs. 75-89.

PENELAS LEGUÍA, A. *Estrategia Comercial en el Sistema Bancario Español: Factores determinates y tipología.* Tesis Doctoral. Universidad de Alcalá, 1994.

SAUNDERS, J.: «Cluster analysis for market segmentation». *European Journal of Marketing* 14, 1980, págs. 422-435.

SYSTAT: The System for Statistics. Evanstons, Il., SYSTAT Inc., 1990

[8] ATM: Activos Totales Medios.

LECTURA COMPLEMENTARIA

El análisis cluster en investigación de marketing: Metodología y crítica

MÓNICA GÓMEZ SUÁREZ
Universidad Autónoma de Madrid[1]

Aceptado para su publicación en *ESIC-Market*

En este trabajo se realiza un análisis crítico de la literatura sobre la aplicación de la técnica de análisis cluster en marketing durante los años noventa. Para ello, se exponen en primer lugar una serie de pasos esenciales a seguir según los expertos cuando se aplica dicha técnica. A continuación se lleva a cabo con posterioridad una análisis exhaustivo sobre la utilización del cluster en marketing en nuestro país (1990-1997).

Palabras clave: *análisis multivariante, análisis cluster, marketing.*

1. INTRODUCCION

El desarrollo informático ha extendido la aplicación en el área de marketing de distintas técnicas de análisis de datos con gran complejidad estadística. Entre los procedimientos más empleados se encuentran el análisis factorial, la regresión múltiple, el análisis discriminante múltiple y el análisis cluster. El objetivo principal de esta última técnica multivariante es la búsqueda de grupos u objetos relativamente homogéneos. Su nombre se utiliza para definir una serie de técnicas que tienen por objeto la búsqueda de grupos similares de individuos o de variables que se van agrupando en conglomerados. Si tenemos una muestra de individuos, de cada uno de los cuales se dispone de una serie de observaciones, el análisis cluster sirve para clasificarlos en grupos lo más homogéneos posible en base a las variables observadas[2]. El origen de esta técnica multivariante se encuentra en la biología y la botánica, debido a que los investigadores de estas áreas de conocimiento tenían la necesidad de agrupar las distintas especies de animales y vegetales conocidas, en familias que fueran lo más homogéneas posibles (Miquel, Bigné y otros, 1997, p. 268). La variedad de nombres con los que se denomina a esta técnica se debe precisamente a su utilización en muy diversas disciplinas, tales como la psicología, la sociología, la economía o la ingeniería. A pesar de ello, todos los métodos desarrollados tienen una dimensión común: clasificar de acuerdo con la relación natural que existe entre los sujetos u objetos (Hair, Anderson y otros, 1995). El análisis cluster se conoce desde hace varias décadas. En los años 30 se desarrollaron algunas tímidas aplicaciones, aunque comenzó a adquirir verdadero auge en los años 50. Las aplicaciones del análisis cluster al marketing han sido múltiples, aunque, lógicamente, el campo donde mayores niveles de aplicación alcanza y donde más beneficios ofrece es el de la segmentación de mercados. También se ha utilizado en la identificación de hábitos de compra, de grupos de productos competitivos, valoración de productos de competidores o identificación de localidades que serán usadas posteriormente como mercados de prueba. En los últimos años, se ha revelado como la técnica más utilizada en el análisis de grupos estratégicos y en el desarrollo de tipologías estratégicas, casi siempre en combinación con otras técnicas multivariantes, como el análisis factorial y el análisis de la varianza. Sin embargo, el uso del análisis cluster también ha sido objeto de numerosos ataques, tanto dentro de la investigación de marketing como de la dirección estratégica (Green, Frank y Robinson, 1967; Morrison, 1967; Schuman, 1967; Funkhouser, 1983; Ketchen y Shook, 1996; Barney y Hoskisson, 1990; Meyer, 1991), fundamentalmente debido a la excesiva confianza que se deposita

[1] Dpto. Financiación e Investigación Comercial. Facultad de CC. Empresariales. UAM. 28049 CANTOBLANCO. MADRID. Tfno: (91)3974348.

[2] La palabra «cluster», que define estas técnicas, se podría traducir como grupo, conglomerado, racimo, apiñarse. En general, en todos los paquetes estadísticos se conserva su nombre inglés, aunque también se conoce como análisis de conglomerados, taxonomía numérica, análisis tipológico o clasificación automática. El cluster se puede utilizar de dos formas muy distintas: clasificación o representación de estructuras de datos. El primer objetivo —clasificación— es la más común de las aplicaciones (Punj y Stewart, 1983). Por esta razón, el artículo se centra en el cluster como forma de clasificación.

en el juicio del investigador al aplicar esta técnica, ya que, a diferencia de otros métodos, el cluster no ofrece un contraste estadístico que facilite una respuesta concreta sobre la confirmación o no de los resultados que sirven de apoyo a las hipótesis planteadas en la investigación y a que muchas aplicaciones del análisis cluster no se han basado en una racionalidad teórica. A menudo, las dimensiones o variables parecen haberse seleccionado al azar y los grupos que se identifican no reflejan condiciones reales sino que son simples «artificios estadísticos supeditados a variaciones numéricas aleatorias entre organizaciones» (Thomas y Venkatraman, 1988). De ahí que con la utilización del cluster se asuma un notable grado de riesgo de ofrecer clasificaciones inexactas, sino imponiendo grupos que realmente no existen.

Incluso algunos autores han afirmado que esta técnica es un «estigma metodológico» (Meyer, 1991), cuestionando su valor como método aplicado a la investigación, debido a los resultados equívocos a los que lleva a veces su utilización (por ejemplo, en la configuración de grupos estratégicos, donde no se ha demostrado de forma consistente la relación entre la pertenencia a un grupo y los resultados). A pesar de esta controversia, el análisis cluster sigue siendo una de las técnicas multivariantes más utililizadas por los investigadores de marketing en nuestro país. Por esta razón, el propósito de este artículo es analizar su aplicación actual dentro de esta disciplina en España, haciendo primero un breve repaso de la situación a nivel internacional. Nuestro objetivo no es sólo evaluar su uso en el pasado, sino también configurar unas líneas maestras que guíen la investigación mediante su aplicación en el futuro, para poder hacer posible su mejor utilización. Para ello, en este trabajo, describimos que aspectos son críticos en la metodología de una investigación cuando se utiliza el cluster. A continuación, realizamos una revisión metodológica en profundidad de los artículos y ponencias publicados en España en los años 90 en las principales revistas y congresos relacionados con esta disciplina. Por último, ofrecemos una serie de recomendaciones que ayuden a los investigadores de marketing a maximizar el valor del análisis cluster. El objetivo de este artículo no es analizar, sin embargo, las primeras aplicaciones en el ámbito del marketing ni hacer un resumen de su gestación como técnica, sino estudiar la utilización en su época de consolidación.

2. ASPECTOS CONCEPTUALES Y CRÍTICOS

En este epígrafe exponemos los principales aspectos relacionados con la aplicación del análisis cluster. No tratamos de ofrecer, sin embargo, una explicación detallada de todos los puntos específicos que acompañan a la aplicación de esta técnica (algoritmos, medidas de distancia, etc), puesto que éstos ya aparecen en una serie de publicaciones anteriores[3], tanto dentro como fuera del área de marketing. No resulta muy difícil agrupar estos aspectos, ya que la mayoría de los expertos han destacado varios de ellos. Sí es una tarea más complicada determinar cuál debe ser la metodología correcta pues, en la mayoría de los casos, estos mismos expertos no han llegado a un acuerdo.

Como primera medida, hemos determinado una serie de puntos esenciales por los que cualquier investigador debe optar al aplicar esta técnica. Al no existir un consenso sobre qué elección es la mejor en algunas ocasiones, el investigador debe analizar las fuerzas o debilidades de un determinado enfoque y elegir áquel que resulte más conveniente a la hora de aportar objetividad a la investigación que desarrolla. Estos puntos esenciales al aplicar el análisis cluster son: la selección de variables, los algoritmos o métodos de agrupación, la determinación del número de grupos y la validación de los resultados[4].

Selección de variables

La selección de variables que se incluyen en el análisis es una paso crucial, ya que si se excluyen variables importantes éste será insuficiente y pobre. En el cluster, la selección inicial de variables determina las características que se usarán para identificar los grupos. Existen fundamentalmente tres métodos de selección de variables: inductivo, deductivo y cognitivo. El primero, en el que ni las variables ni el número de grupos que se espera obtener tienen un nexo con una teoría, es el que más se utiliza en estudios exploratorios. El segundo se apoya en la literatura teórica preexistente. A diferencia del anterior, la consistencia interna de los grupos es mayor, ya que no se incluyen variables irrelevantes que disminuyen la validez de las agrupaciones resultantes. El método congnitivo está muy cercano, en la práctica, al inductivo, ya que no se basa en apreciaciones teóricas (hipótesis), sino que utiliza las predicciones de determinados expertos de la industria para definir las variables (Ketchen y Shook, 1996). La selección de variables basada en uno de estos tres métodos depende de los objetivos de la investigación. Si existe una teoría previa, se debe confirmar mediante la utilización de las variables que facilita. No obstante, en muchas ocasiones, los estudios realizados en el campo del marketing son exploratorios y, por tanto, se debe recurrir a un enfoque inductivo o cognitivo. Si tenemos acceso a las opiniones de expertos, es recomendable considerarlas, ya que de esta forma aumentaremos la probabilidad de que las variables reflejen fielmente la realidad.

La estandarización de las variables y la utilización de factores que provienen de un análisis factorial previo son también dos aspectos que se deben tratar en este epígrafe.

[3] Por ejemplo, Klastorin (1983), Punj y Stewart (1983), Bisquerra (1989), Noururis (1993), Hair y otros (1995), Kinnear y Taylor (1996), Miquel, Bigné y otros (1997).

[4] El esquema que seguimos se basa en el artículo publicado por Ketchen y Shook en el Strategic Management Journal en 1996. Dicho artículo analizaba todos las investigaciones sobre estrategia publicadas en las seis revistas prestigiosas de management (Academy of Management Journal, Journal of International Business Studies, Journal of Management, Management Science y Strategic Management Journal) desde 1977 hasta 1993.

Al utilizar diferentes unidades de medida en las distintas variables, áquellas que se midan con grandes números solaparán a las que se miden con números pequeños. Para resolver este problema se pueden estandarizar las variables aunque no siempre es la mejor estrategia, ya que el proceso de estandarización permite que las variables contribuyan de igual forma a la definición de los grupos, pero también puede eliminar diferencias significativas entre los elementos de análisis (Funkhouser, 1983). Debido a que no hay un consenso entre los expertos en cuanto a la estandarización o no de las variables, la mejor estrategia es emplear ambos enfoques (estandarización y no estandarización). Si la solución no es consistente, entonces se debe mirar la validez de los resultados de cada uno de ellos y elegir áquel cuya solución aporte mayor confianza. En cuanto a la utilización de factores que provienen de una análisis factorial previo, se pretende eliminar con ella la multicolinealidad entre variables. También la utilización de factores, en vez de variables, ha suscitado controversias entre los investigadores, ya que al emplear sólo los factores que tienen mayor autovalor se puede eliminar información que pueda ser importante a la hora de diferenciar entre los individuos[5]. En este caso, ocurre como con la estandarización, se debe determinar el impacto que tiene la utilización de los factores o de las variables por separado en la consistencia de los grupos que se obtienen mediante ambos enfoques.

Algoritmos o métodos de agrupación

Existen muchas formas de combinar los objetos. Ninguno de los métodos proporciona una solución óptima a todos los problemas, ya que es posible llegar a distintos resultados según el método elegido por el investigador. En general, se suele distinguir entre métodos jerárquicos y no jerárquicos (también denominados iterativos). Los primeros van agrupando o dividiendo los grupos sin tener que determinarse a priori el número de grupos final. Estos procedimientos se clasifican, a su vez, en aglomerativos (se agrupan los casos en grupos hasta que se forma un único grupo) y en divisivos (parte del grupo total y lo corta en grupos hasta llegar a cada caso concreto). Si se utilizan métodos no jerárquicos, se debe dar una solución a priori en cuanto al número de grupos que se forman. El primer paso del método aglomerativo consiste en considerar cada caso como un grupo distinto. Después se combinan dos de los casos en un sólo grupo, un tercer caso se añade al cluster o se combina con otro para formar un cluster distinto. Así pues, en cada paso, un caso individual se añade a los clusters ya existentes o se combina con otro para formar un nuevo cluster. Una vez formado el cluster, ya no se puede cortar, sólo se puede combinar con otros clusters. En general hay tres tipos de submétodos dentro del aglomerativo: Métodos de encadenamiento, Métodos de varianza y Métodos centroides.

Los métodos jerárquicos divisivos se han utilizado menos en las ciencias sociales, fundamentalmente por la inexistencia de programas que permitieran realizar el análisis de los datos. Entre los algoritmos más conocidos se puede citar el de Howard-Harris, que forma grupos, por división de otros de tamaño mayor, de modo también secuencial, utilizando el criterio de minimización de la varianza intragrupos en cada nivel de la división (Santesmases, 1997). Las principales desventajas asociadas a la utilización de los métodos jerárquicos son la dificultad de determinar a priori el mejor algoritmo cuando el investigador desconoce la estructura de la muestra y la inestabilidad de las soluciones que se aprecia cuando el tamaño muestral es reducido y/o hay presencia de casos atípico (outliers). En cuanto a los métodos no jerárquicos, el más utilizado es el k-means. Su principal inconveniente se deriva de que, a menudo, no se sabe cuál es el número de grupos que debe haber. Sin embargo, las soluciones se ven menos afectadas por la existencia de casos atípicos. La solución que suelen dar los expertos en cuanto a la elección del método de agrupación es la de combinar ambos tipos de algoritmos en dos etapas: primero determinar el número de grupos mediante un método jerárquico y, una vez definida la mejor solución en cuanto a número de grupos, utilizar un algoritmo no jerárquico[6]. Asimismo, se debe considerar la medida de distancia o similitud utilizada[7].

Determinación del número de grupos

La literatura ofrece diferentes criterios que pueden ayudar al investigador a tomar la decisión relativa al número óptimo de grupos a formar. Cuando se aplica un método jerárquico, el criterio más frecuentemente utilizado es la observación del dendrograma (representación gráfica del proceso de agrupación). También se acude a la observación del coeficiente de aglomeración (valor numérico en el que se une varios casos que forman un nuevo grupo), bien mediante gráficos (coeficiente en un eje y número de casos en otro), bien mediante la utilización de medidas matemáticas que expresan los cambios incrementales en el coeficiente[8]. Otra forma de obtener los grupos se consigue mediante el ajuste global obtenido con cada grupo y la mejora experimentada en dicho

[5] Por ejemplo, Dillon, Mullani y Frederick (1989) demuestran que al aplicar un análisis de componentes principales y elegir sólo aquellos factores que explican mayor varianza, se eliminan dos factores que sirven para determinar las mayores diferencias entre los estilos de vida de distintos grupos de personas pertenecientes a la tercera edad.

[6] Si se desea ver una comparación entre los métodos de agrupación jerárquicos y no jerárquicos se puede consultar a Bigné, Miquel y otros (1997, pág. 276).

[7] Véase anexo 2.

[8] Las medidas que se suelen utilizar se basan en la homogeneidad dentro del grupo en relación a la heterogeneidad entre grupos. Existen hasta treinta índices distintos desarrollados por distintos expertos, siendo los más eficientes el de Calinski y Harabatz, el Je(2)/Je(1), el índice C y el índice de Beale (Milligan y Cooper, 1985).

ajuste con un grupo adicional[9]. Por último, la existencia de una teoría previa puede facilitar la determinación de los grupos sin basarse en una medida gráfica o estadística.

Validación de los grupos

En este caso, se trata de determinar tanto la fiabilidad como la validez de las soluciones que se han alcanzado. En cuanto a la fiabilidad, se debe probar que los resultados son consistentes, utilizando distintos algoritmos de agrupación y distintos métodos que muestren la inexistencia de multicolinealidad. Asimismo, se puede demostrar la fiabilidad partiendo la muestra de tal forma que una mitad sirva para definir los grupos y la otra para probar su consistencia. Sin embargo, muchas veces es difícil acudir a esta medida porque las muestras son muy pequeñas. Si se ha demostrado la fiabilidad, aún queda por probar la validez, tanto externa (la solución sobre la muestra es representativa de la población) como interna (es útil para predecir resultados). En cuanto a la validez externa, se debe acudir a una muestra parecida de la población que determine los mismos resultados. Respecto a la interna, se deben utilizar contrastes estadísticos que permitan establecer la consistencia de la solución. Normalmente se acude al análisis de la varianza o a otra técnica multivariante como el análisis discriminante[10].

A lo largo del epígrafe hemos puesto de manifiesto que la mayoría de estos aspectos dependen del juicio del investigador, ya que ni los mismos expertos en la técnica se ponen de acuerdo en cuanto a su utilización. No obstante, siempre se pueden emplear métodos alternativos y comprobar su consistencia y validez. En el siguiente epígrafe, expondremos la metodología seguida para practicar el estudio de los procedimientos más extendidos en la aplicación del análisis cluster al maketing, tanto a nivel internacional, como en España.

3. PROCESO DE REVISIÓN DE LA LITERATURA

3.1. Metodología

La revisión de la literatura acerca de la aplicación del cluster al marketing se centra en una revisión profunda de la bibliografía publicada en España, limitándonos al período 1990-1997, ya que la mayoría de los trabajos han ido apareciendo desde principios de la década actual.

En cuanto a los trabajos a nivel nacional, al no existir una base de datos completa sobre revistas españolas, hemos revisado exclusivamente las publicaciones más frecuentes sobre temas de marketing, concretamente, *ESIC Market, Distribución y Consumo, Revista Europea de Dirección y Economía de la Empresa, Información Comercial Española, Investigación y Marketing y Revista Española de Investigación de Marketing.* Debemos reconocer, por tanto, el sesgo que hemos introducido en esta investigación al no tener en cuenta otras publicaciones. La otra fuente bibliográfica de obligada referencia son las actas de los encuentros de Profesores de Marketing. Respecto a las tesis doctorales, consultamos la base de datos TESEO.

Las primeras revisiones que se realizaron en el área de Marketing sobre la utilización del análisis cluster se basaron en la construcción estadística de los algoritmos de agrupación, llegando a concluir que algunos de ellos eran bastante limitados (Sherman y Seth, 1977; Frank y Green, 1968). Posteriormente, las revisiones obviaron la parte más estadística y se centraron en la aplicación de la técnica. Una de las revisiones metodológicas más consistentes es la de Punj y Stewart (1983), cuyo objetivo consiste en clarificar el empleo del cluster, facilitando unas recomendaciones específicas al usuario de la técnica. Concretamente, la revisión abarca la naturaleza de los datos, el algoritmo utilizado y su descripción, así como los principales programas de ordenador que permiten aplicar el cluster[11].

3.2. Resultados de la revisión bibliográfica en España

Con la revisión que hemos realizado en España tratamos de demostrar cómo se está aplicando la técnica del cluster a partir de los puntos que hemos considerado clave, ya descritos

[9] El criterio de parada es el siguiente (Flavián y Polo, 1997, pág.110):

1. El porcentaje de la varianza intragrupo explicado con cada agrupación debe superar el 60 por ciento.
2. El incremento del porcentaje en la explicación de la varianza intra con la inclusión de un nuevo grupo en cada etapa debe ser superior al 5 por ciento.

[10] Existen otras técnicas aparte del anova o discriminante como los que sugieren Klastorin (1983) o Punj y Stewart (1983), aunque estas dos técnicas son las más utilizadas.

[11] Se ha efectuado anteriormente una revisión bibliográfica a nivel internacional que se apoya en las bases de datos en formato CD-ROM, en concreto ABI-INFORM, ECONLIT y UMI (tesis doctorales), siguiendo la metodología de palabras clave con el fin de determinar cuantas publicaciones y cómo ha sido la evolución numérica de aplicación del análisis cluster. Desde mediados de los años ochenta (primeros artículos que aparecen en las base de datos) hasta la actualidad se han publicado 77 artículos en las principales revistas de marketing, fundamentalmente en el *European Journal of Marketing,* en el *Journal of the Academy of Marketing Science* (con 11 en cada una de las revistas), en el *Journal of Advertising Research* (10) y en el *Journal of Marketing Research* (8). Asimismo, las tesis doctorales que se han valido de la técnica de cluster superan el centenar (106 tesis). La conclusión más relevante cuando se analiza el cuadro es que la aplicación de la técnica objeto de nuestro estudio disfrutó de un período de introducción bastante largo, fundamentalmente debido a la incorporación paulatina de programas de software que permitían realizar el análisis cluster y al acceso de investigadores a esos programas estadísticos, siendo su aplicación notable a finales de los ochenta y principios de los 90. Desde 1994 los investigadores han ido reduciendo su utilización. Nos debemos preguntar porqué el empleo de esta técnica ha disminuido de forma tan significativa.

en el Apartado 2. En concreto, estos aspectos son el método de determinación de variables, su estandarización y su agrupación en factores, el algoritmo de agrupación, la forma de determinación del número de grupos, la validación de los resultados, el programa de ordenador utilizado y el tema de marketing en el que se pueden encuadrar las investigaciones. Los resultados de esta revisión se pueden observar en el Tabla 9.L.1.

TABLA 9.L.1. Resumen de los aspectos relacionados con la aplicación del análisis cluster

	N.º Estudios	*Porcentaje*
1. Selección de variables		
1.1. Método		
Inductivo	8	15,6
Deductivo	18	35,2
Inductivo/Cognitivo	25	49,0
1.2. Análisis factorial previo	20	39,2
1.3. Estandarización de las variables	9	17,6
2. Algoritmos de agrupación		
2.1. Jerárquicos	10	19,6
Completo	4	40,0
Ward	2	20,0
Howard-Harris	2	20,0
Entre grupos	1	10,0
Otros	1	10,0
2.2. No jerárquicos	22	43,1
2.3. Combinación	10	23,8
2.4. No especifican	24	47,1
3. Determinación del número de grupos		
3.1. Múltiples métodos	7	13,7
3.2. Método simple	11	21,6
3.3. Otros procedimientos	9	17,6
3.4. No especifican	24	47,1
4. Validación de los grupos		
4.1. Fiabilidad	9	17,6
4.2. Validez		
No especifican	21	41,2
Discriminante	11	21,6
Anova-Manova	17	33,3
Discriminante/Anova	2	3,9
5. Programa estadístico		
5.1. SPSS	33	64,7
5.2. BMDP	1	2,0
5.3. Galaxy	3	5,9
5.4. Dyane	2	3,9
5.5. Systab	1	2,0
5.6. Spad	3	5,9
5.7. No especifican	8	15,7
6. Tipos de análisis		
6.1. Grupos estratégicos	11	21,6
6.2 Segmentación	17	33,3
6.3. Otros	23	45,1
7. Tema de marketing		
7.1. Comportamiento del consumidor	9	18,0
7.2. Comunicación	4	8,0
7.3. Distribución	12	24,0
7.4. Producto-servicio	19	38,0
7.5. Marketing de servicios	5	10,0
7.6. Otros	5	10,0

Elaboración propia basada en la revisión de los artículos publicados en las revistas.

Respecto a la **selección de variables**, el método más utilizado ha sido una combinación entre inductivo y cognitivo (25 estudios que suponen el 49 por ciento del total). También se acude con frecuencia al método deductivo (35 por ciento). Este resultado pone de manifiesto que aunque casi la mitad de los estudios tienen un carácter exploratorio, la teorías existentes sobre el tema que se está investigando son abundantes y permiten seleccionar las variables que sirven para configurar los análisis. Resulta llamativo el dato sobre el bajo número de estudios que estandarizan las variables (9 de 51, lo que supone un 17 por ciento del total). Asimismo, 20 (39 por ciento) de los 51 estudios examinados han procedido a realizar un factorial previo, utilizando las cargas como input de información para el cluster. Tanto la estandarización como la utilización de las cargas factoriales tienen detractores y defensores entre los expertos. Sin embargo, es importante que los investigadores comprueben la influencia de la multicolinealidad, así como el solapamiento de las variables con unidades de medida grandes. Por tanto, nosotros recomendamos que se prueben las distintas soluciones que se obtienen al aplicar ambas medidas, puesto que demostrarán la consistencia de la solución. Sobre todo, en cuanto a la correlación existente entre las variables es conveniente que se indique en los artículos alguna medida que demuestre si ésta puede influir en la agrupación. En cuanto a **los algoritmos de agrupación** se observa que el método más preferido ha sido el no jerárquico (43 por ciento) frente a la combinación de métodos (24 por ciento) o los métodos jerárquicos (19 por ciento). Debemos destacar además que un 17 por ciento de los estudios no especifican qué procedimiento de agrupación se ha utilizado.

Creemos que la preferencia por el método no jerárquico (fundamentalmente K-means), se debe a que en su momento la capacidad de memoria de los ordenadores no facilitaba el uso de un método alternativo, sobre todo con grandes archivos, de tal forma que los investigadores sólo podían aplicar esta técnica. De hecho, se puede ver que al avanzar en el tiempo se ha ido sustituyendo la utilización de k-means por métodos combinados. La combinación preferida ha sido la aplicación conjunta del método de Ward con k-means, determinando con el primer algoritmo el número de grupos y empleando después los grupos obtenidos con el método no jerárquico. Respecto al método de Ward se debe advertir que tiende a sobrevalorar a los casos atípicos (outliers), lo cual suele influir en la solución final. Por ello, y como en el caso anterior, nuestra recomendación es que se utilicen varios métodos que permitan demostrar la validez de la solución obtenida. En este sentido, la mejor estrategia consiste en seguir un procedimiento basado en dos etapas (Punj y Stewart, 1983). En la primera se puede utilizar un método jerárquico que nos ayude a decidir el número de grupos a identificar, los centroides y los casos atípicos que deben ser eliminados. Los casos restantes serán objeto de iteraciones partitivas mediante un método no jerárquico aplicado con posterioridad (véase Figura 9.L.1). La **determinación del número de grupos** es quizá uno de los puntos más débiles de todas las investigaciones realizadas, puesto que el 47 por ciento de los estudios no especifican cómo han llegado a obtener los grupos. Los que sí lo hacen utilizan un sólo método, normalmente la observación del dendrograma, con lo que las agrupaciones que se realizan tienen una alta dosis de subjetividad, dependiendo en gran medida del juicio del investigador. Creemos que la utilización de distintos procedimientos puede ayudar a determinar cuál es el número de grupos idóneo. Asimismo, estimamos conveniente que en el futuro los investigadores especifiquen tanto el algoritmo de agrupación como la forma en que han determinado el número de clusters.

También las indicaciones sobre la **validación de la técnica** es un aspecto a resaltar en esta revisión, ya que sólo el 17 por ciento de los estudios muestran la fiabilidad, existiendo además 21 investigaciones (41 por ciento) que no se preocupan de demostrar la validez. Quizá resulta difícil probar la fiabilidad a través de cortes en la muestra, ya que muchas veces estas muestras son pequeñas, pero sí se pueden aplicar distintos métodos de agrupación que nos ayuden a mostrar la fiabilidad. Respecto a la validez externa ningún estudio ha tratado de generalizar los datos a la población, por las dificultades que este procedimiento conlleva. No obstante, creemos que no es tan difícil probar la validez de criterio mediante la aplicación de contrastes estadísticos u otras técnicas multivariantes que muestren la consistencia de los resultados. De las 30 investigaciones que prueban dicha validez, la técnica más utilizada es el análisis de la varianza (33 por ciento), seguida del análisis discriminante (21 por ciento) o la combinación de ambas (4 por ciento). La utilización o preferencia por determinados algoritmos de agrupación está condicionada por los programas de ordenador con los que se analiza la información. El programa más usado es el SPSS (64 por ciento), seguido muy de lejos por el SPAD y el Galaxy (3 investigaciones cada uno). Ocho investigaciones no especifican el paquete estadístico. Estos trabajos tampoco mencionan el algoritmo de agrupación, debiendo el lector confiar en el juicio del autor sobre las decisiones tomadas al aplicar esta técnica y no pudiendo replicar posteriormente los resultados de la investigación. Por último, hemos tratado de sintetizar en la Tabla 9.L.1 los temas prioritarios en las investigaciones que han empleado el cluster. Como primer dato importante, destacamos que el 33 por ciento de los estudios se ha centrado en segmentación y el 21 por ciento en grupos estratégicos, sobre todo, dentro del sector bancario. Desagregando por temas, el marketing de servicios (38 por ciento), la distribución (24 por ciento) y el comportamiento del consumidor (18 por ciento) son las áreas de aplicación donde esta técnica se ha desarrollado con más profusión.

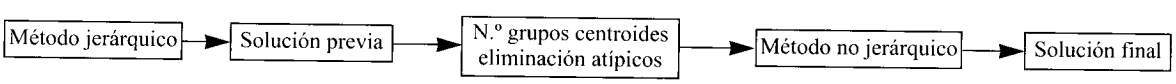


FIGURA 9.L.1. Proceso de elección del número de grupos.

5. CONCLUSIONES

El análisis cluster es un método estadístico de clasificación que no necesita acudir a supuestos previos sobre la pertenencia a un grupo por parte de la población. Es un método empírico puro, por lo que puede ayudar al desarrollo de la ciencia del marketing cuando el objetivo primario de la investigación es clasificar consumidores, productos u organizaciones. No obstante, esta técnica debe aplicarse con prudencia para que su validez sea plena. En concreto, su utilización tiene que ceñirse a unas normas que se basan, sobre todo, en el uso conjunto de varios métodos alternativos y de apoyo. En este artículo hemos puesto de manifiesto que muchas de las investigaciones realizadas en España hasta el momento no han cumplido parte de los requisitos inherentes a la aplicación del cluster. Muchas de ellas no mencionan los algoritmos de agrupación o cómo se determina el número de grupos. Asimismo, no se prueba, en la mayoría de los casos, la fiabilidad y la validez de las soluciones. Esta falta de especificación o de prueba impide la repetición de los resultados en investigaciones posteriores y ofrece poca ayuda respecto al procedimiento a utilizar por otros investigadores (Punj y Stewart, 1983). El trabajo aquí presentado pretende facilitar la tarea de los investigadores que se plantean utilizar el cluster por primera vez o ayudar a perfeccionar su uso en el caso de que se haya aplicado con anterioridad, puesto que pese a sus inconvenientes es la única técnica que nos permite clasificar a priori una serie de objetos o sujetos, siendo la clasificación el primer y último paso de la investigación científica.

BIBLIOGRAFÍA

AGUDO, A. y FERNÁNDEZ, J. M. *Métodos de segmentación basados en la aplicación combinada de técnicas estadísticas*, VI Encuentro de Profesores Universitarios de Marketing, San Sebastián, 1994, pp. 99-107.

BARNEY, J. B. y HOSKINSON, R. E. «Strategic groups: untested assertions and research proposals», *Managerial Decision Economics,* vol. 11, pp. 187-198, 1990.

BEERLI, A. y GARCÍA, J. M. «La segmentación del mercado y su aplicación al mercado bancario minorista», *Información Comercial Española,* n.º 707, 1992, pp. 143-151.

BEERLI, A. y RODRÍGUEZ, A. *El comportamiento del mercado de productos de confección textil: hacia una tipología de sus consumidores,* VI Encuentro de Profesores Universitarios de Marketing, San Sebastián, 1994, pp. 382-394.

BIGNE, E. «El análisis cluster: una aplicación a las áreas de interés en de las agencias de publicidad», *Esic-Market,* n.º 69, 1990, pp. 7-28.

BIGNE, E., VALLET, T., MOLINER, M. y SÁNCHEZ, J. *La satisfacción de usuarios de servicios públicos hospitalarios,* VIII Encuentro de Profesores Universitarios de Marketing, Zaragoza, 1996, pp. 231-240.

BISQUERRA, R. *Introducción conceptual al análisis multivariable: un enfoque infromático con los paquetes SPSS-X, BMDP, LISREL y SPAD*, vol. II, Barcelona, Promociones y Publicaciones Universitarias, 1989.

CALOMARDE, J. V. *El análisis cluster aplicado a la segmentación del comportamiento de compra por variables ecológicas,* III Encuentro de Profesores Universitarios de Marketing. Salamanca, 1991, pp. 317-345.

CALORMARDE, J. V. (1992). *Actitudes ante el medio ambiente y centros de interés social,* IV Encuentro de Profesores Universitarios de Marketing, El Escorial, 1992.

CALORMARDE, J. V. «Actitudes ante el medio ambiente y centros de interés social», *Esic-Market,* n.º 89, 1995, pp. 125-154.

CALOMARDE, J. V., GARCÍA, B. y MARTÍNEZ, E. *La influencia de los factores ecológicos en las decisiones de marketing empresarial: el sector del papel y del cartón,* IX Encuentro de Profesores Universitarios de Marketing, Murcia, 1997, pp. 43-64.

CAMARERO, C. y MARTÍN, N. *La comunicación ecológica de las empresas: un estudio empírico,* VII Encuentro de Profesores Universitarios de Marketing, Barcelona, 1995, pp. 99-109.

CAMARERO, C., GUTIÉRREZ, I. y RODRÍGUEZ, I. *Insatisfacción, comportamiento de queja e intención de recompra: un estudio exploratorio (el caso de la educación universitaria),* VIII Encuentro de Profesores Universitarios de Marketing, Zaragoza, 1996, pp. 43-54.

CÉSPEDES, J. J. y SÁNCHEZ, M. «Tendencias y desarrollos recientes en métodos de investigación y análisis de datos en dirección de empresas», *Revista Europea de Dirección y Economía de la Empresa,* vol. 5, n.º 3, 1996, pp. 23-40.

DILLON, W. R., MULANI, N. y FREDERICK, D. G. «On the use of componenet scores in the presence of group structure», *Journal of Consumer Research,* vol. 16, 1989, pp. 106-112.

DÍAZ, A., RÍO, B., SANTOS, L. y SANZO, M. J. *Los valores sociales del consumidor: un estudio para el mercado del automóvil,* VIII Encuentro de Profesores Universitarios de Marketing, Zaragoza, 1996, pp. 30-41.

ESTEBAN, A., GARCÍA, J. y NARROS, M. J. *Análisis del comportamiento de la demanda turística española procedente de Europa,* VI Encuentro de Profesores Universitarios de Marketing, San Sebastián, 1994, pp. 470-478.

ESTEBAN, A., GARCÍA, J. y NARROS, M. J. «Análisis del comportamiento de la demanda turística española procedente de Europa», *Esic-Market,* n.º 85, 1994, pp. 183-197.

ESTEBAN, A., GARCÍA, J. y NARROS, M. J. *Análisis comparativo del comportamiento: la demanda turística del primer trimestre de 1994-1995,* VII Encuentro de Profesores Universitarios de Marketing, Barcelona, 1995, pp. 503-512.

FLAVIÁN, C. y POLO, Y. «Los grupos estratégicos como herramienta de análisis de la competencia en el sector detallista español», *Esic-Market,* n.º 96, 1997, pp. 9-28.

FLAVIÁN, C. y POLO, Y. «Identificación de patrones de comportamiento estratégico en la gran distribución española», *Información Comercial Española,* n.º 763, 1997. pp. 99-117.

FONT, B. «Los procedimientos de análisis cluster aplicados a la distribución comercial», *Esic-Market,* n.º 94, 1996, pp. 73-83.

FRASQUET, M. y MOLLA, A. «Variedad en el comportamiento de la clientela en centros comerciales», *Información Comercial Española,* n.º 763, 1997, pp. 138-150.

FUENTES, F., BARRIO, S. y SÁNCHEZ, J. *Importancia de los atributos calidad/precio en el proceso de decisión del consumidor granadino,* IX Encuentro de Profesores Universitarios de Marketing, Murcia, 1997, pp. 171-183.

FUNKHOUSER, G. R. «A note on the reability of certain clustering algorithms», *Journal OF Marketing Research,* vol. 20, 1983, pp. 99-102.

GARCÍA, J.M. y RODRÍGUEZ, M. «Grupos estratégicos en el sector del comercio textil: una metodología fundamental en las variables del retailing-mix», *Información Comercial Española,* n.º 750, 1996, pp. 133-152.

GÓMEZ, M. *Estrategias de distribución en banca de empresas,* IX Encuentros de Profesores Universitarios de Marketing, Murcia, 1997, pp. 185-201.

GÓMEZ, M. y MÉNDEZ, J. L. *Imagen y posicionamiento de entidades financieras en la CAM,* V Encuentro de Profesores Universitarios de Marketing, Sevilla, 1993, pp. 269-279.

GRANDE, I. y ABASCAL, E. *Criterios de elección de establecimiento comercial: el caso de la tercera edad,* V Encuentro de Profesores Universitarios de Marketing, Sevilla, 1993, pp. 171-180.

GREEN, P. E., FRANK, R. E. y ROBINSON, P. J. (1967). «Cluster analysis in test market selection», *Management Science,* vol. 13, 1967, pp. 387-400.

HAIR, J. H., ANDERSON, R. E., TATHAM, R. L. y BLACK,W. C. *Multivariate data analysis with readings,* New Jersey, Prentice Hall, 4.ª edición, 1995.

HERNÁNDEZ, M., MUNUERA, J. L. y RUIZ DE MAYA, S. «La estrategia de diferenciación en el comercio minorista», *Información Comercial Española,* n.º 739, 1995, pp. 27-45.

IGLESIAS, V. «Tipos de variables y emtodología a emplear en la identificación de grupos estratégicos: una aplicación empírica al sector detallista», *Revista Europea de Dirección y Economía de la Empresa,* vol. 3, n.º 3, 1995, pp. 73-86.

IGLESIAS, V. y RUIZ, A. «El análisis de las estrategias de marketing en el comercio minorista», *Información Comercial Española,* n.º 763.

JIMÉNEZ, A. y VARGAS, M. *El efecto de la satisfacción en la lealtad de la marca,* VII Encuentro de Profesores Universitarios de Marketing, Barcelona, 1995, pp. 271-281.

KLASTORIN, T. D. «Assesing cluster analysis resoults», *Journal of Marketing Research,* vol. XX, 1983, pp. 92-98.

KETCHEN, D. J. y SHOOK, C. L. «The application of cluster analysis in strategic management research: an analysis and critique», *Strategic Management Journal,* vol. 17, 1996, pp. 441-458.

LUQUE, T. y CORDON, E. «Una aplicación del análisis multivariable a las características socio-económicas y comerciales de las capitales de provincia españolas», *Revista Europea de Dirección y Economía de la Empresa,* vol. 3, n.º 1, 1994, pp. 101-112.

LUQUE, T., FRIAS, D. y MARAVER, G. *Las escuelas de marketing y los encuentros de profesores universitarios,* VI Encuentro de Profesores Universitarios de Marketing, San Sebastián, 1994, pp. 13-25.

LUQUE, T., SÁNCHEZ, G. y FRÍAS, D. «La innovación en el mercado farmaceútico español», *Revista Europea de Dirección y Economía de la Empresa,* vol. 4, n.º 1, 1995, pp. 47-58.

MAS RUIZ, F. *Análisis de la competencia en Cajas de Ahorros: grupos estratégicos,* III Encuentro de Profesores Universitarios de Marketing. Salamanca, 1991, pp. 260-263.

MAS RUIZ, F. y GÓMEZ SALA, J. «Identificación de los grupos estratégicos en las Cajas de Ahorros españolas», *Revista Europea de Dirección y Economía de la Empresa,* vol. 1, n.º 3, 1992, pp. 9-28.

MAS RUIZ, F. y GOMEZ SALA, J. «Análisis de la competencia en cajas de ahorros: grupos estratégicos», *Esic-Market,* n.º 79, 1993, pp. 143-167.

MÉNDEZ, J. L. «La actitud de los consumidores frente a las formas comerciales: resultados de una investigación sobre hábitos de compra», *Distribución y Consumo,* n.º 4, 1992, pp. 47-53.

MEYER, A. D. «What is strategy's distinctive competence?», *Academy of Management Journal,* vol. 36, 1991, pp. 1175-1195.

MILLIGAN, G. W. y COOPER, M. C. «An examination of procedures for determining the number of clusters in a data set», *Psychometrika,* vol. 50, n.º 2, 1985, pp. 59-179.

MIQUEL, S. y BIGNE, E. *Un análisis cluster del sector hotelero español,* IV Encuentro de Profesores Universitarios de Marketing, El Escorial, 1992.

MIQUEL, S., BIGNE, E., LEVY, J. P., CUENCA, A. y MIQUEL, M. J. *Investigación de mercados,* Madrid, McGraw-Hill, 1997.

MOLINER, M. A., VALLET, T. y SÁNCHEZ, J. *Percepción y posicionamiento de los servicios públicos. Una aplicación a la ciudad de Castellón,* VI Encuentro de Profesores Universitarios de Marketing, San Sebastián, 1994, pp. 153-163.

MORRISON, D. G. «Measurement problems in cluster analysis», *Management Science,* vol. 13, 1967, pp. 775-780.

MUNUERA, J. L., RUIZ DE MAYA, S., FERNÁNDEZ, M. y MÁS RUIZ, F. *Información Comercial Española,* n.º 718, 1995, pp. 119-138.

MUÑIZ, N. *E volución de los grupos estratégicos en la distribución comercial europea,* VIII Encuentro de Profesores Universitarios de Marketing, Zaragoza, 1996, pp. 326-338.

MUÑOZ, P. «Estrategias de posicionamiento en las empresas de distribución comercial», *Distribución y Consumo,* n.º 10, 1993, pp. 46-57.

NARROS, M. J. «Métodos de segmentación de mercado: un estudio comparativo aplicado a las audiencia de radio», *Esic-Market,* n.º 87, 1994, pp. 53-67.

ONIS, T. y RIALP, J. *La elección de una universidad: el*

caso concreto de la UAB, VIII Encuentro de Profesores Universitarios de Marketing, Zaragoza, 1996, pp. 93-102.

PENELAS, A. *Generación de estrategias naturales en el marketing bancario*, IV Encuentro de Profesores Universitarios de Marketing, 1992, El Escorial.

PUNJ, G. y STEWART, D. W. «Cluster analysis in Marketing Research. Review and suggestions for application», *Journal of Marketing Research,* vol. 20, 1983, pp. 134-148.

RODRÍGUEZ, I., TRESPALACIOS, J. y VÁZQUEZ, R. *La actitud como determinante del grado de satisfacción de un servicio,* VI Encuentro de Profesores Universitarios de Marketing, San Sebastián, 1994, pp. 413-423.

RODRÍGUEZ, I., TRESPALACIOS, J. Y VÁZQUEZ, R. «La actitud como determinante del grado de satisfacción de un servicio», *Esic-Market,* n.º 85, 1994, pp. 157-170.

RUIZ, A. *Segmentación del mercado financiero de economías familiares: un estudio empírico*, V Encuentro de Profesores Universitarios de Marketing, Sevilla, 1993, pp. 291-306.

RUIZ, A. «Segmentación del mercado financiero de economías familiares: un estudio empírico», *Esic-Market,* n.º 97, 1994, pp. 9-41.

RUIZ, A., GONZÁLEZ, F. e IGLESIAS, V. *Comercialización de un producto indiferenciado: la gasolina en España,* VIII Encuentro de Profesores Universitarios de Marketing, Zaragoza, 1996, pp. 311-325.

SANTESMASES, M. *Dyane: diseño y análisis de encuestas en investigación social y de mercados.* Madrid, Pirámide, 1997.

SCHUMAN, A. «Letter to the editor», *Management Science,* vol. 13, 1967, pp. 688-696.

THOMAS, H. y VENKATRAMAN, N. «Research on strategic groups: progress and prognosis», *Journal of Management Studies,* vol. 25, 1988, pp. 537-555.

TRESPALACIOS, J., RÍO, A. e IGLESIAS, V. *Dimensiones de la orientación al mercado y sus efectos sobre las empresas del sector agroalimentario,* IX Encuentros de Profesores Universitarios de Marketing, Murcia, 1997, pp. 428-443.

VÁZQUEZ, R., RODRÍGUEZ, I. y RUIZ, A. *Calidad de servicio y su percepción por el consumidor: aplicación a las empresas detallistas,* VII Encuentro de Profesores Universitarios de Marketing, Barcelona, 1995, pp. 415-431.

CAPÍTULO 10

Análisis multidimensional

OBJETIVOS DE APRENDIZAJE

- Apreciar cómo la representación espacial de los datos puede clarificar las relaciones subyacentes entre sujetos u objetos.
- Determinar el número de dimensiones representadas en los datos.
- Interpretar mapas espaciales de tal forma que se puedan apreciar las dimensiones.
- Entender las diferencias entre los datos de similitud y los datos de preferencia tal y como se utilizan en el análisis multidimensional.
- Distinguir entre las diferentes aproximaciones de aplicación del análisis multidimensional.

El análisis multidimensional (MDS) consiste en una serie de técnicas que ayudan al investigador a identificar las dimensiones subyacentes claves en las evaluaciones de los objetos de estudio por parte de los encuestados. Por ejemplo, el análisis multidimensional se utiliza a menudo en marketing para identificar las dimensiones subyacentes determinantes en las evaluaciones de los productos, servicios o compañías por parte de los clientes. Otras aplicaciones habituales incluyen la comparación de calidades físicas (por ejemplo, gustos alimenticios o diversos olores), percepciones de asuntos o candidatos políticos, e incluso la evaluación de diferencias culturales entre diferentes grupos. Las técnicas de análisis multidimensional pueden inferir las dimensiones subyacentes de una serie de juicios de preferencia o similitud realizados por los encuestados sobre los objetos de estudio. Una vez que se dispone de los datos, el análisis multidimensional puede ayudar a determinar (1) qué dimensiones utilizan los encuestados cuando evalúan los objetos, (2) cuántas dimensiones pueden utilizarse en una situación particular, (3) la importancia relativa de cada dimensión y (4) cómo se relacionan perceptualmente los objetos.

¿Qué es el análisis multidimensional?

El análisis multidimensional, también conocido como elaboración de mapas perceptuales, es un procedimiento que permite al investigador determinar la imagen percibida relativa de un conjunto de objetos (empresas, productos, ideas u otros asuntos asociados con percepciones habituales). El propósito del análisis multidimensional es transformar los juicios del consumidor de similitud o preferencia (es decir, preferencia por establecimientos o marcas) en distancias representadas en un espacio multidimensional. Suponga que se juzga a los objetos A y B por parte de encuestados, que los consideran lo más similares comparados con todos los posibles pares de objetos. Las técnicas de análisis multidimensional situarán los objetos A y B de tal forma que la distancia entre ellos en el espacio multidimensional es menor que la distancia entre cualquier otro par de objetos. El **mapa perceptual** resultante, también conocido como **mapa espacial**, muestra la situación relativa de todos los objetos, tal y como se muestra en la Figura 10.1.

El análisis multidimensional se basa en las comparaciones entre **objetos**. Se puede suponer que cualquier objeto (por ejemplo, producto, servicio, imagen, aroma) tiene *dimensiones* objetivas y percibidas. Por ejemplo, la dirección de HATCO puede considerar su producto (un cortacésped) con dos opciones de color (rojo y verde), un motor de dos caballos de potencia y una hoja de 24 pulgadas. Éstas son **dimensiones objetivas**. Por otro lado, los clientes pueden (o no) observar estos atributos. Los clientes pueden percibir el cortacésped de HATCO como caro o frágil. Éstas son **dimensiones percibidas**, también conocidas como **dimensiones subjetivas**. Dos productos pueden tener las mismas características físicas (dimensiones objetivas), pero ser vistos de forma diferente, debido a que las distintas marcas se perciben de forma diferente, en calidad (una dimensión percibida) por los clientes. Por tanto, son muy importantes las dos diferencias entre dimensiones objetivas y perceptuales que se señalan a continuación:

1. *Diferencias individuales*: las dimensiones percibidas por los clientes pueden no coincidir con (o pueden no incluir) las dimensiones objetivas supuestas por el investigador. Esperamos que cada individuo pueda tener diferentes dimensiones percibidas, pero el investigador debe también aceptar que las dimensiones objetivas pueden variar sustancialmente. Los individuos pueden considerar diferentes conjuntos de características objetivas así como variar la importancia que dan a cada dimensión.
2. *Interdependencia*: las evaluaciones de las dimensiones (incluso si las dimensiones percibidas son las mismas que las dimensiones objetivas) pueden no ser independientes y pue-

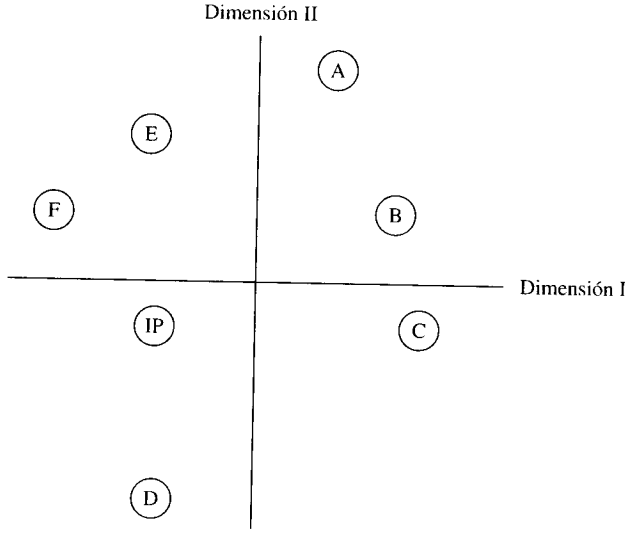


Figura 10.1. Ilustración de un mapa multidimensional de percepciones de seis proveedores industriales (A hasta F) y el punto ideal (IP).

den no concordar. Tanto las dimensiones objetivas como las percibidas pueden interactuar con el resto creando evaluaciones inesperadas. Por ejemplo, un refresco puede juzgarse más dulce que otro, debido a que el primero tiene un aroma más parecido a la fruta, aunque contenga la misma cantidad de azúcar.

El desafío para el investigador está, en primer lugar, en entender las dimensiones percibidas y a continuación referirlas a dimensiones objetivas, si es posible. Se necesita un análisis adicional para evaluar qué atributos predicen la posición de cada objeto tanto en el espacio objetivo como en el perceptual.

Debe recomendarse precaución, sin embargo, en la interpretación de las dimensiones. Debido a que este proceso es más un arte que una ciencia, el investigador debe resistir la tentación de permitir que su percepción personal afecte a la dimensionalidad cualitativa de las dimensiones percibidas. Dado el nivel de la opinión del investigador, se deben tomar todas las precauciones con el fin de ser lo más objetivo posible en esta crítica, y todavía rudimentaria, área.

Una visión simplificada del análisis multidimensional

Para facilitar una mejor comprensión de los procedimientos básicos del análisis multidimensional, presentaremos en primer lugar un ejemplo simple con el fin de ilustrar los conceptos básicos subyacentes al análisis multidimensional y el procedimiento por el cual transforma juicios de similitud en las correspondientes posiciones espaciales.

Los investigadores de mercado están interesados en comprender las percepciones de los clientes de seis chocolatinas presentes en el mercado. En lugar de intentar conseguir información sobre las evaluaciones de los consumidores sobre las chocolatinas a través de un cierto número de atributos, los investigadores reunirán sólo percepciones de similitudes y diferencias conjuntas. Los datos se reúnen normalmente pidiendo a los encuestados respuestas globales a cuestiones como éstas:

- Calificación de similitud de los productos A y B en una escala de 1 a 10.
- El producto A es más similar a B que a C.
- Prefiero el producto A al B.

De estas respuestas simples, se puede deducir el mapa perceptual que representa mejor la pauta global de similitudes entre las seis chocolatinas. Ilustraremos el proceso de creación de un mapa perceptual con los datos de un único encuestado, aunque este proceso podría aplicarse a varios encuestados o a las respuestas agregadas de un grupo de consumidores.

Los datos se consiguen creando en primer lugar un conjunto de 15 pares de chocolatinas ($6 \times 5/2 = 15$ pares). A los encuestados se les pregunta a continuación que califiquen los siguientes 15 pares de chocolatinas, donde una calificación de 1 se da al par de chocolatinas que es más parecido y la calificación 15 indica que el par es el menos parecido. En la Tabla 10.1 se muestran los resultados (órdenes de calificación) de todos los pares de chocolatinas de un encuestado.

El encuestado cuyas calificaciones se muestran en la Tabla 10.1 piensa que las chocolatinas D y E son las más parecidas, las chocolatinas A y B serían las siguientes más parecidas y así sucesivamente hasta las chocolatinas E y F, que son las menos parecidas. Si queremos ilustrar la similitud entre las chocolatinas gráficamente, en un primer intento se podría deducir una única **escala de similitud** y ajustar todas las chocolatinas a esta escala. Esto implica una representación unidimensional de similitud, donde la distancia representa la similitud. Por tanto, los objetos cercanos en la escala son más parecidos y aquellos más alejados son menos parecidos. El objetivo es situar las chocolatinas en la escala de tal forma que las órdenes de calificación estén mejor representadas (la calificación 1 es la más cercana, la calificación 2 es la siguiente más próxima, y así sucesivamente).

TABLA 10.1. Datos de similitud (órdenes de calificación) para pares de chocolatinas.

Chocolatina	*A*	*B*	*C*	*D*	*E*	*F*
A	—	2	13	4	3	8
B		—	12	6	5	7
C			—	9	10	11
D				—	1	14
E					—	15
F						—

Nota: los valores más bajos indican mayor similitud; 1 indica los pares más parecidos y 15 los pares menos parecidos.

Intentemos ver cómo colocaríamos algunos de los objetos. Colocar dos o tres chocolatinas es bastante simple; la primera prueba real llega con cuatro objetos. Elegimos las chocolatinas A, B, C y D. La Tabla 10.1 muestra que el rango de orden de los pares es el siguiente: $\overline{AB} < \overline{AD} < \overline{BD} < \overline{CD} < \overline{BC} < \overline{AC}$ (la línea que hay sobre cada par de letras indica que la expresión se refiere a la distancia [similitud] entre los pares). A partir de estos valores, debemos situar las cuatro chocolatinas en una escala única, de tal forma que las más similares ($\overline{AB}$) sean las más cercanas y las menos similares ($\overline{AC}$) sean las más alejadas. La Figura 10.2 (parte A) contiene un mapa perceptual unidimensional que ajusta el orden de los pares descritos más arriba. Si la persona que juzga la similitud entre las chocolatinas ha estado pensando en términos de una regla simple de similitud que contiene sólo un atributo (dimensión), tal como la cantidad de chocolate, entonces todos los pares podrían ser situados en una escala simple que reproduce los valores similares.

Aunque se puede realizar un mapa unidimensional con cuatro objetos, la tarea se hace cada vez más difícil a medida que el número de objetos aumenta. Se anima al lector interesado a intentar la tarea con seis objetos. Cuando se emplea una única dimensión con seis objetos, el orden efectivo varía sustancialmente con respecto a las calificaciones originales del encuestado. Dado que un análisis unidimensional no se ajusta bien a los datos, se debería intentar una solución bidimensional. Esto permitiría utilizar otra dimensión en la configuración de las seis chocolatinas. Intentar a mano este procedimiento es bastante tedioso; en la Figura 10.2 (parte B) se muestra la solución bi-dimensional producida por un programa de análisis multidimensional. Esta configuración se ajusta exactamente a las calificaciones de la Tabla 10.1, apoyando la noción de que el encuestado utilizará probablemente dos dimensiones para evaluar las chocolatinas. La conjetura de al menos dos atributos (dimensiones) se considera que está basada en la incapacidad de representar las percepciones de un encuestado en una dimensión. Sin embargo, todavía no somos conscientes de qué atributos utiliza el encuestado en la evaluación.

Aunque no tenemos información acerca de cuáles son estas dimensiones, podemos ser capaces de observar las posiciones relativas de las chocolatinas e inferir qué atributos representan las dimensiones. Por ejemplo, suponga que las chocolatinas A, B y F fueran una forma de combinación de la chocolatina (por ejemplo, chocolate y cacahuetes, chocolate y mantequilla de cacahuetes) y C, D y E fueran estrictamente de chocolate. Podríamos inferir a continuación que la dimensión I representa el tipo de chocolatina (chocolate frente a combinación). Cuando observamos la posición de las chocolatinas en la dimensión vertical, pueden emerger otros atributos como descriptores de la dimensión.

El análisis multidimensional permite a los investigadores entender las similitudes entre las seis chocolatinas preguntando sólo por las percepciones de similitud. El procedimiento puede ayudar también a determinar qué atributos forman parte en realidad de las percepciones de similitud. Aunque no incorporamos directamente las evaluaciones de atributos en el procedimiento de análisis multidimensional, podemos utilizarlos en análisis subsecuentes para ayudar a interpretar las dimensiones y los impactos que cada atributo tiene en las posiciones relativas de las chocolatinas.

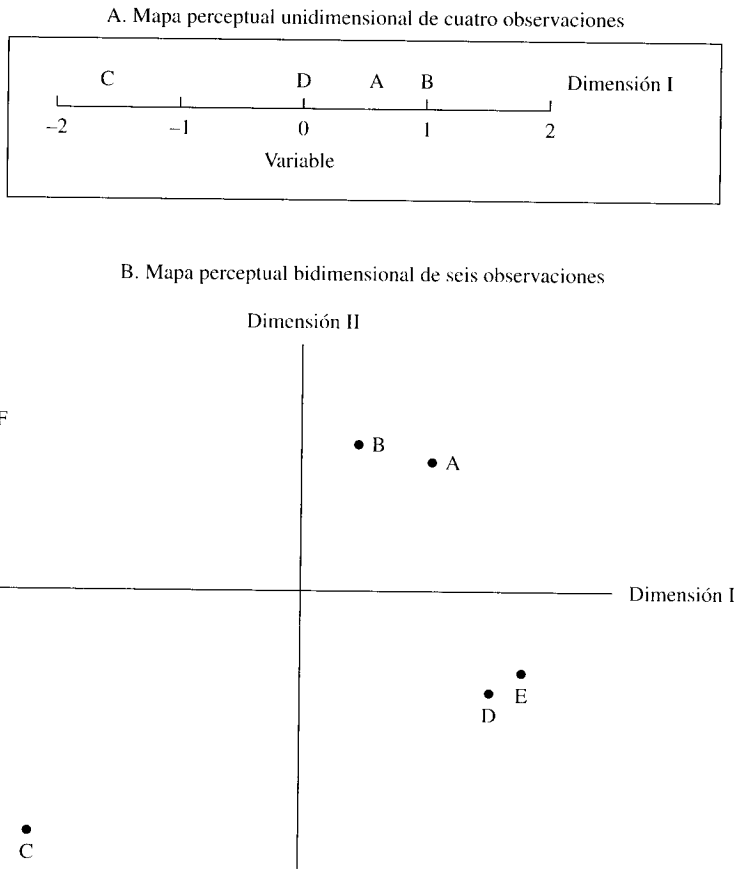


FIGURA 10.2. Mapa perceptual de una y dos dimensiones.

Comparación del análisis multidimensional con otras técnicas de interdependencia

El análisis multidimensional puede compararse con otras técnicas de interdependencia, tales como el análisis factorial y el análisis cluster en función de su aproximación a la estructura definida. El análisis factorial agrupa variables en valores teóricos que definen las dimensiones subyacentes del conjunto original de variables. Las variables que están muy correlacionadas se agrupan conjuntamente. El análisis cluster agrupa observaciones de acuerdo con su perfil sobre un conjunto de variables (valor teórico cluster) en el cual las observaciones muy cercanas se agrupan juntas. El análisis multidimensional difiere del análisis cluster en dos aspectos claves: (1) se puede obtener una solución para cada individuo, y (2) no utiliza el valor teórico.

El individuo como unidad de análisis

En el análisis multidimensional, cada encuestado proporciona evaluaciones de todos los objetos que se están considerando, por lo que se puede obtener una solución para cada individuo, lo que no es posible para el análisis cluster o factorial. Como tal, el énfasis no se pone en los objetos mismos, sino en cómo el individuo percibe los objetos. La estructura que está siendo definida se basa en las dimensiones perceptuales de comparación de los individuos. Una vez que las dimensiones perceptuales están definidas, se pueden realizar las comparaciones relativas entre los objetos.

Ausencia de valor teórico

El análisis multidimensional, a diferencia de otras técnicas multivariantes, no utiliza el valor teórico. Por el contrario, las «variables» que compensarían al valor teórico (es decir, las dimensiones perceptuales de comparación) se infieren de medidas globales de similitud entre objetos. Haciendo una analogía simple, esto es como proporcionar la variable dependiente (similitud entre objetos) e imaginarse cuál debe ser la variable independiente (dimensiones perceptuales). El análisis multidimensional tiene la ventaja de reducir la influencia del investigador al no exigir la especificación de las variables que se utilizan en la comparación de objetos, tal y como se requiere en el análisis cluster. Pero también tiene la desventaja de que el investigador no está realmente seguro de qué variables está utilizando el encuestado para realizar las comparaciones.

Un marco de decisión para elaborar mapas perceptuales

La elaboración de mapas perceptuales abarca un amplio rango de métodos posibles, incluyendo el análisis multidimensional, y todas estas técnicas pueden verse a través del proceso de construcción de modelos como el introducido en el Capítulo 1. Estos pasos representan un marco de decisión, mostrado en la Figura 10.3 (niveles 1-3) y 10,4 (niveles 4-6), dentro del cual pueden aplicarse todas las técnicas de elaboración de mapas perceptuales y sus resultados pueden ser evaluados.

Primer paso: Objetivos del análisis multidimensional

La elaboración de mapas perceptuales, y el análisis multidimensional en particular, son los más apropiados para conseguir dos objetivos:

1. Como una técnica exploratoria para identificar dimensiones no reconocidas que afectan al comportamiento.
2. Como medio de obtener evaluaciones comparativas de objetos cuando las bases específicas de comparación no se conocen o no están definidas.

En el análisis multidimensional, no es necesario especificar los atributos de comparación de los encuestados. Todo lo que se requiere es especificar los objetos y asegurarse de que comparten una base común de comparación. Esta flexibilidad hace al análisis multidimensional particularmente apropiado para realizar estudios gráficos y de situación en los que las dimensiones de evaluación sean muy globales o muy emocionales y afectivas como para ser medidas por análisis convencionales. Algunos métodos de análisis multidimensional combinan la situación de objetos y sujetos en un único mapa conjunto. En estas técnicas, las posiciones relativas de los objetos y consumidores hacen el análisis de segmentación mucho más directo.

Decisiones clave en la fijación de objetivos

Una característica común a cada objetivo es la falta de especificación en la definición de los niveles de evaluación de los objetos. El punto fuerte de la elaboración de mapas perceptuales está en su habilidad para «inferir» dimensiones sin la necesidad de definir atributos. La flexibilidad y la naturaleza inferencial del análisis multidimensional exige al investigador una mayor responsa-

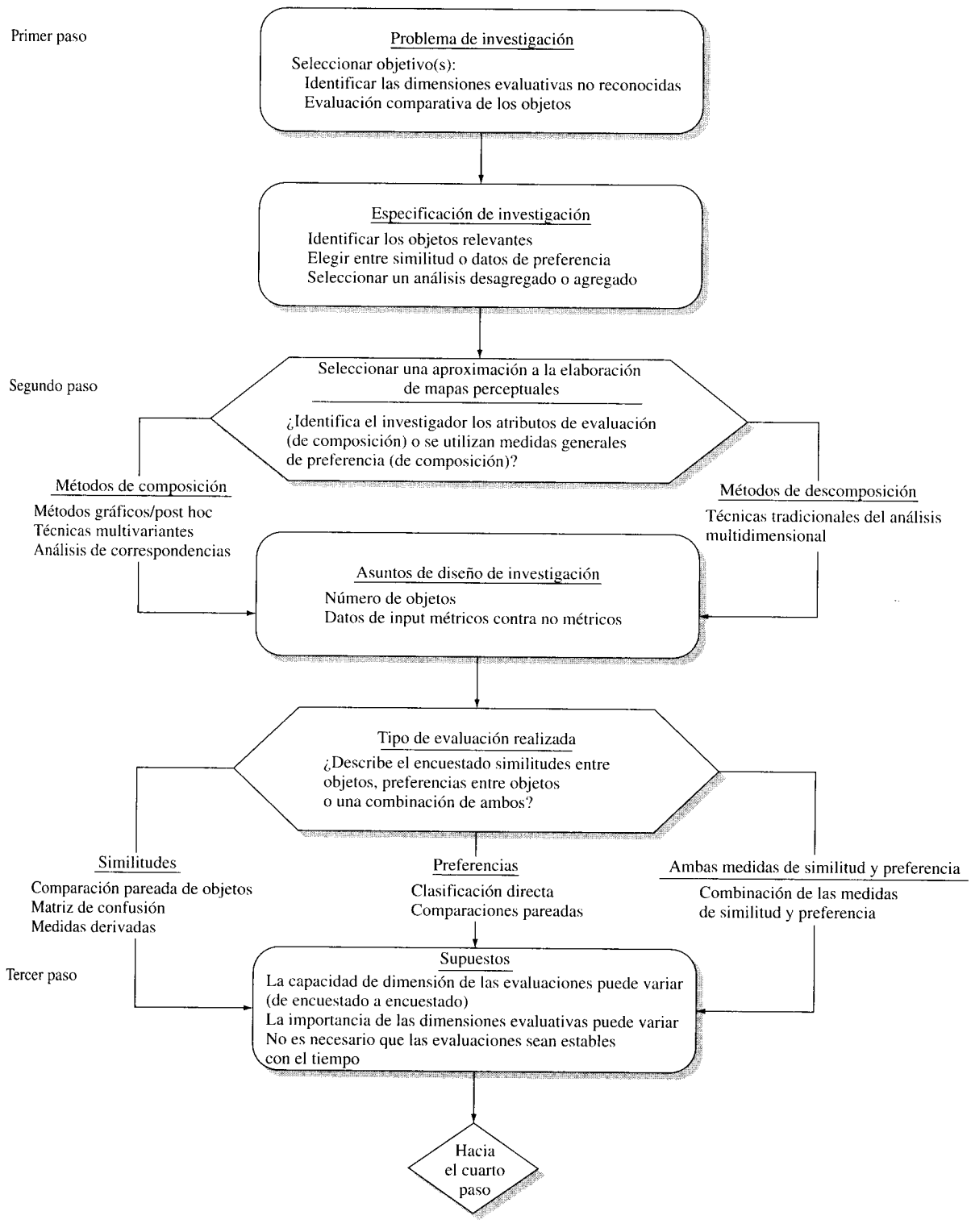


FIGURA 10.3. Pasos 1-3 en el diagrama de decisión del análisis multidimensional.

bilidad a la hora de definir «correctamente» el análisis. Tanto las consideraciones prácticas como las conceptuales son esenciales para que el análisis multidimensional consiga sus mejores resultados. Para asegurar este éxito, el investigador debe definir un análisis multidimensional a través de tres decisiones clave: seleccionar los objetos que se van a evaluar, decidir si son las similitudes o las preferencias lo que se va analizar y elegir si el análisis será realizado a nivel de grupo o individual.

Identificación de todos los objetos relevantes que se van a evaluar

El asunto más básico, y a la vez importante, en la elaboración de mapas perceptuales es la definición de los objetos a evaluar. El investigador debe asegurarse que todas las empresas, productos, servicios u otros objetos «relevantes» están incluidos, dado que el mapa perceptual es una técnica de situación relativa. La relevancia está determinada por las preguntas que hace el investigador. Los mapas perceptuales resultantes de cualquier método pueden verse enormemente influenciados tanto por la omisión como por la inclusión de objetos inapropiados [7, 20]. Si se incluyen los «objetos» irrelevantes o no comparables, el investigador está forzando la técnica no sólo para inferir las dimensiones perceptuales que distinguen entre objetos comparables, sino también para inferir aquellas dimensiones que distinguen entre objetos no comparables. Esta tarea está más allá del alcance del análisis multidimensional y resulta en una solución que no concuerda con la realidad.

Datos de similitud frente a datos de preferencia

Una vez seleccionados los objetos de estudio, el investigador debe elegir a continuación las bases de evaluación: similitud frente a preferencia. Hasta este punto, hemos discutido la elaboración de mapas perceptuales y análisis dimensional en términos de juicios de similitud. Al proporcionar **datos de similitud**, el encuestado no aplica ningún aspecto de evaluación en términos de «bueno-malo» en la comparación. La evaluación «bueno-malo» se hace, por el contrario, dentro de los datos de **preferencia**, que asumen que diferentes combinaciones de atributos percibidos se valoran más que otras combinaciones. Ambas bases de comparación se pueden utilizar para desarrollar mapas perceptuales, pero con diferentes interpretaciones. Los mapas perceptuales basados en la similitud representan similitudes de atributos y dimensiones perceptuales de comparación pero no reflejan ninguna visión directa en los determinantes de la elección. Los mapas perceptuales basados en la preferencia reflejan elecciones preferidas pero pueden no corresponder de ningún modo a posiciones basadas en la similitud, dado que los encuestados pueden basar sus elecciones en dimensiones o criterios enteramente diferentes de aquellos en los que basan sus comparaciones. No existe una base óptima de evaluación, pero la decisión entre datos de similitudes y de preferencia debe realizarse teniendo en cuenta la investigación a realizar, dado que son fundamentalmente diferentes en lo que representan.

Análisis agregado frente a análisis desagregado

Al considerar datos de preferencia o similitud, estamos tratando las percepciones de estímulos de los encuestados y creando representaciones de proximidad entre estímulos en un espacio t-dimensional (donde el número de dimensiones t es menor que el número de estímulos). El investigador puede generar este resultado sujeto a sujeto (produciendo tantos mapas como sujetos), un método conocido como **análisis desagregado**. Una de las características distintivas de las técnicas de análisis multidimensional es su capacidad para estimar soluciones para cada encuestado, que pueden ser representadas a continuación por separado. La ventaja es la representación de los elementos genuinos de cada percepción del encuestado. La desventaja es que el investigador debe identificar los elementos comunes a todos los encuestados.

Las técnicas de análisis multidimensional también pueden combinar encuestados y crear menos mapas perceptuales mediante un proceso denominado **análisis agregado**. La agregación pue-

de realizarse tanto antes como después del análisis de los datos del encuestado. Antes del análisis, la aproximación más simple para el investigador es encontrar las evaluaciones «medias» para todos los encuestados y obtener una solución simple para el grupo de encuestados en conjunto. Para identificar grupos de encuestados similares, el investigador puede analizar mediante el método cluster las respuestas de los encuestados para encontrar un número reducido de encuestados o representativos o medios y a continuación desarrollar mapas para cada individuo y agrupar los mapas de acuerdo con las coordenadas de los estímulos en los mapas. Es más recomendable utilizar las evaluaciones «medias» en lugar de agregar los mapas de percepción individual, debido a que el menor número de rotaciones del mismo mapa, puede crear problemas al obtener aglomerados razonables mediante el segundo método.

Se puede encontrar una forma especializada de análisis desagregado en INDSCAL (análisis de diferencias individuales) [4] y sus variantes, que tienen las características de ambos tipos de análisis. INDSCAL supone que todos los individuos comparten un espacio común o de grupo (una solución agregada) pero que todos los encuestados individualmente ponderan las dimensiones, incluyendo las ponderaciones cero cuando ignoran totalmente una dimensión. Como primer paso, INDSCAL obtiene el mapa perceptual compartido por todos los individuos, de la misma forma que hacen otras soluciones agregadas. Sin embargo, los individuos también están retratados en un mapa perceptual de un grupo especial. Aquí, la posición de los encuestados se determina por sus ponderaciones de cada dimensión. Los encuestados situados cerca emplean similares combinaciones de dimensiones para formar el espacio común de grupo. Más aún, la distancia de los individuos desde el origen es una medida aproximada de la proporción de la variación de ese sujeto tenida en cuenta en la solución. Por tanto, una posición alejada del origen indica un buen ajuste. Estar en el origen significa «no ajuste», debido a que todas las ponderaciones son cero. Si dos o más sujetos o grupos de sujetos están en el origen, se necesitan configurar otros espacios de grupo para cada uno de ellos. En este análisis, el investigador tiene presente no sólo una representación del mapa perceptual, sino también el grado en que está representado cada individuo por el mapa perceptual. Estos resultados de cada encuestado pueden utilizarse a continuación para agrupar a los encuestados e incluso identificar diferentes mapas perceptuales en análisis subsecuentes.

La elección de análisis agregados o desagregados se basa en los objetivos de estudio. Si el objetivo es un entendimiento de las evaluaciones conjuntas de objetos y las dimensiones empleadas en esas evaluaciones, es más apropiado un análisis agregado. Pero si el objetivo es entender la variación entre individuos, entonces es mejor utilizar el análisis desagregado.

Segundo paso: Diseño de investigación mediante el análisis multidimensional

Aunque el análisis multidimensional parece bastante simple a efectos de cálculo, los resultados, como ocurre con otras técnicas multivariantes, están poderosamente influidos por un número de problemas clave que deben resolverse antes de que pueda empezar la investigación. Cubrimos cuatro de los problemas principales, que van desde la discusión del diseño de la investigación (seleccionando la aproximación y los objetos o estímulos de estudio) a la especificación de asuntos metodológicos (métodos métricos frente a los no métricos) y métodos de recogida de datos.

Selección de un enfoque de descomposición (libre de atributos) o de composición (basada en atributos)

Las técnicas de elaboración de mapas perceptuales pueden clasificarse en función de la naturaleza de las respuestas obtenidas del individuo en relación con el objeto. Una aproximación, el **método de descomposición**, mide sólo la impresión o evaluación conjunta de un objeto y a continuación intenta obtener posiciones espaciales en un espacio multidimensional que refleje es-

tas percepciones. Esta técnica se asocia típicamente con el análisis multidimensional. El **método de composición** es una aproximación alternativa, que emplea varias técnicas multivariantes ya discutidas que se usan en la formación de una impresión o evaluación basada en una combinación de atributos específicos. Cada aproximación tiene ventajas y desventajas. Nuestra discusión aquí se centra en las distinciones entre las dos aproximaciones y después nos centraremos principalmente en los métodos de descomposición.

Enfoque de descomposición o libre de atributos

Habitualmente asociado con las técnicas de análisis multidimensional, los métodos de descomposición se basan en medidas conjuntas o globales de similitud, de las que se forman los mapas perceptuales y el posicionamiento relativo de los objetos. Tiene dos ventajas. La primera es que sólo requiere que los encuestados expresen percepciones conjuntas de los objetos; los encuestados no detallan los atributos utilizados en esta evaluación. La segunda es que, dado que cada encuestado da una evaluación de similitudes completa entre todos los objetos, se pueden desarrollar mapas perceptuales para encuestados individuales o para conjuntos de éstos para formar un mapa de composición.

Los métodos de descomposición tienen también desventajas. La primera, que el investigador no tiene una base objetiva provista por el encuestado sobre la cual identificar las «dimensiones» básicas de evaluación de los objetos (es decir, la correspondencia de dimensiones objetivas y perceptuales). En muchos ejemplos, la utilidad de los estudios libres de atributos está restringida porque los estudios arrojan poca luz para una acción específica. Por ejemplo, la incapacidad de desarrollar un vínculo directo entre acciones de la empresa (la dimensión objetiva) y las posiciones de mercado de sus productos (la dimensión perceptual) muchas veces disminuye el valor de la elaboración de mapas perceptuales. Además, el investigador tiene poca ayuda, aparte de líneas generales y creencias a priori, en la determinación tanto de la dimensionalidad del mapa perceptual como de la representatividad de la solución. Aunque existen algunas medidas globales de ajuste, no son estadísticas, y por tanto las decisiones sobre la solución final implican una interpretación por parte del investigador.

Existe un amplio rango de técnicas de descomposición, caracterizadas bajo la categoría general de técnicas de análisis multidimensional. La selección de un método específico requiere decisiones en relación con la naturaleza de las respuestas de los encuestados (calificación o clasificación), si se obtienen similitudes o diferencias y si se utilizan mapas perceptuales individuales o compuestos. Entre los programas más habituales de análisis multidimensional están KYST, MDSCAL, PREFMAP, MDPREF, INDSCAL, ALSCAL, MINISSA, POLYCON y MULTISCALE. En Schiffman *et. al.* [23, 24] se pueden encontrar descripciones detalladas de los programas y las formas de obtenerlos.

Enfoque de composición o basado en atributos

Los métodos de composición incluyen algunas de las técnicas multivariantes más habituales (por ejemplo, el análisis discriminante o el análisis factorial), así como métodos diseñados específicamente para la elaboración de mapas perceptuales, tales como el análisis de correspondencias. No obstante, un principio común para todos estos métodos es la evaluación de similitud en la cual se considera un conjunto de atributos en el desarrollo de la similitud entre objetos.

Una ventaja de esta aproximación es la descripción explícita de las dimensiones del espacio perceptual. Debido a que el encuestado proporciona evaluaciones detalladas de numerosos atributos acerca de cada objeto, los criterios de evaluación representados por las dimensiones de la solución son más fáciles de averiguar. En segundo lugar, estos métodos ofrecen un método directo de representación de ambos (atributos y objetos) en un mapa único, donde diversos métodos ofrecen un posicionamiento adicional de los grupos de encuestados. Esta información supone una visión única para los responsables empresariales en un mercado competitivo.

Hay cuatro desventajas fundamentales de las técnicas de composición. En primer lugar, la similitud entre objetos está limitada sólo a atributos calificados por los encuestados. Si se omiten los

atributos más destacados, no hay oportunidad por parte del encuestado para introducirlos, como ocurre si se ofrece una única medida global. En segundo lugar, el investigador debe *asumir* algún método de combinación de estos atributos para representar la similitud conjunta, y el método elegido puede no representar el pensamiento de los encuestados. En tercer lugar, el esfuerzo en la recolección de los datos es importante, especialmente a medida que el número de objetos a elegir aumenta. Por último, los resultados habitualmente no están a disposición del encuestado individual.

Los métodos de composición pueden agruparse en tres grupos básicos:

1. *Enfoques básicos o* post hoc—se incluyen en este conjunto análisis tales como gráficos de diferencia semántica o **parrillas de importancia-realización**, que descansan en el juicio del investigador y representaciones univariantes o bivariantes de los objetos.
2. *Técnicas estadísticas multivariantes convencionales*—estas técnicas, especialmente el *análisis factorial* y *análisis discriminante*, son particularmente útiles en el desarrollo de una estructura dimensional entre los numerosos atributos, para representar a continuación objetos sobre estas dimensiones.
3. *Métodos especializados de elaboración de mapas perceptuales*—en esta clase de técnicas destaca el *análisis de correspondencias*, desarrollado específicamente para proporcionar una elaboración de mapas perceptuales con datos analizados sólo cualitativa o nominalmente como input.

Selección entre técnicas de composición y descomposición

La elaboración de mapas perceptuales puede desarrollarse con técnicas de composición o descomposición, pero cada técnica tiene ventajas y desventajas específicas que deben considerarse a la vista de los objetivos de investigación. Si la elaboración de mapas perceptuales forma parte del «espíritu» de uno de los dos objetivos básicos discutidos anteriormente, los enfoques de descomposición o libres de atributos son los más apropiados. No obstante, si, los objetivos de la investigación se encaminan a la representación entre objetos sobre un conjunto definido de atributos, entonces las técnicas de composición son la alternativa preferida. Nuestra discusión de los métodos de composición en los capítulos pasados ha ilustrado sus usos y aplicaciones junto con sus puntos fuertes y sus debilidades. El investigador siempre debe recordar las alternativas que están disponibles en el caso de que los objetivos del investigador cambien. Por tanto, aquí nos centramos en los enfoques de descomposición, seguido por una discusión del análisis de correspondencias, una técnica de composición ampliamente utilizada que es particularmente apropiada para la elaboración de mapas perceptuales. Como tales, consideramos como sinónimos los términos *elaboración de mapas perceptuales* y *análisis multidimensional* a menos que se tengan que hacer distinciones necesarias.

Objetos: Su número y selección

Antes de comenzar cualquier estudio de elaboración de mapas perceptuales, el investigador debe responder varias cuestiones sobre los objetos que se están evaluando. La primera y más importante, ¿son los objetos realmente comparables? Un supuesto implícito en la elaboración de mapas perceptuales es que existen características comunes, tanto objetivas como percibidas, que el encuestado podría utilizar para las evaluaciones. No es posible para el investigador «forzar» al encuestado a comparar creando pares de objetos no comparables. Incluso si las respuestas se diesen en situaciones forzadas, su utilidad es cuestionable.

Una segunda cuestión hace referencia al número de objetos a evaluar. Al decidir cuántos objetos hay que incluir, el investigador debe tener en cuenta dos deseos: un reducido número de objetos que faciliten el esfuerzo por parte del encuestado, frente al número de objetos requerido para obtener una solución multidimensional estable. Una pauta sugerida para las soluciones estables es tener más de cuatro veces tantos objetos como dimensiones deseadas [9]. Por tanto, se requieren

al menos cinco objetos para un mapa perceptual unidimensional, nueve objetos para una solución bidimensional, y así sucesivamente. Cuando se utiliza el método de evaluación de pares de objetos por similitud, el encuestado debe hacer 36 comparaciones de nueve objetos —una tarea importante. Y una solución de tres dimensiones supone que habría que analizar al menos 13 objetos, y por tanto evaluar 78 pares de objetos. En resumen, hay que compensar entre la dimensionalidad adecuada a los objetos (y el número implicado de dimensiones subyacentes que pueden ser identificados) y el esfuerzo exigido por parte del encuestado.

El número de objetos también puede afectar a la determinación de un nivel aceptable de ajuste. Muchas veces, tener menos del número de objetos sugerido para una dimensionalidad dada provoca una estimación inflada del ajuste. De la misma forma que en el problema del sobreajuste que encontramos en la regresión, estar fuera de las líneas recomendadas de cuatro objetos por dimensión aumenta enormemente las posibilidades de obtener una solución equivocada. Por ejemplo, un estudio empírico demuestra que cuando siete objetos se ajustan a tres dimensiones con valores aleatorios de *similitud*, el 50 por ciento de las veces se generan mapas perceptuales válidos y niveles de stress aceptables. Si los siete objetos con similitudes aleatorias se ajustaran a las cuatro dimensiones, los valores de stress caerían a cero, indicando un ajuste perfecto, en la mitad de los casos [18]. Aunque en ambos casos no exista un patrón real de similitud entre los objetos. Por tanto, debemos tener precaución a la hora de incumplir las pautas del número de objetos por dimensión y el impacto que tiene tanto sobre las medidas de ajuste como sobre la validez de los mapas perceptuales resultantes.

Métodos métricos frente a métodos no métricos

Los programas originales de análisis multidimensional eran enteramente no métricos, lo que significa que exigían sólo datos no métricos pero que a su vez proporcionaban sólo resultados no métricos (clasificación-orden). Los resultados no métricos, sin embargo, limitaban la interpretabilidad del mapa perceptual. Sin embargo, todos los mapas utilizados hoy en día producen resultados métricos. Las posiciones métricas multidimensionales pueden girarse desde el origen, el origen puede cambiarse añadiendo una constante, los ejes pueden ser volteados (reflejo), o incluso la solución entera se puede comprimir o encoger de manera uniforme, sin cambiar la posición relativa de los objetos.

Dado que todos los programas actuales producen resultados métricos, la distinción se basa en la *medida de los datos de similitud*. Los métodos no métricos, que se distinguen por el dato no métrico típicamente generado por pares de objetos clasificados según rango, son más flexibles en la medida en que no suponen ningún tipo de relación específica entre la distancia calculada y la medida de similitud. Sin embargo, debido a que los métodos no métricos contienen menos información para crear el mapa perceptual, es más probable que arrojen soluciones subóptimas o degeneradas. Éste es un problema particular cuando existen amplias variaciones en los mapas perceptuales entre los encuestados o las percepciones entre los objetos no se distinguen o no están bien definidas. Los métodos métricos suponen que los datos y los resultados sean métricos. Este supuesto nos permite estrechar la relación entre la dimensionalidad del resultado final y los datos de entrada. En lugar de suponer que sólo las relaciones ordenadas están preservadas en los datos de entrada, podemos asumir que el resultado preserva el intervalo y las calidades del ratio de los inputs de entrada. Incluso aunque los supuestos subyacentes bajo los programas métricos son más difíciles de apoyar conceptualmente en muchos casos, los resultados de procedimientos métricos y no métricos aplicados a los mismos datos son a menudo muy similares. Por tanto, al seleccionar el tipo de datos de entrada debe considerarse tanto la situación de investigación (variaciones de percepciones entre encuestados y distinciones entre los objetos) como el modo preferido de recogida de datos.

Recogida de datos sobre similitudes o preferencias

Como ya se ha apuntado, la distinción principal entre los programas de análisis multidimensional se basa en el tipo de datos (métricos o no métricos) utilizados para representar similitudes o preferencias. Por tanto, preguntamos a los encuestados asuntos asociados con hacer juicios de pre-

ferencia o basados en similitudes. Para muchos de los métodos de obtención de datos, no pueden recogerse ni los datos métricos (calificaciones) ni los no métricos (clasificaciones). En muchos casos, sin embargo, las respuestas están limitadas a un solo tipo de datos.

Datos de similitudes

Cuando se recogen **datos de similitudes**, el investigador está intentando determinar qué puntos son más parecidos a otros y cuáles son los más diferentes. Los términos de similitudes y diferencias a menudo se utilizan indistintamente para representar medidas de las diferencias entre objetos. Implícita en la medida de similitud está la capacidad de comparar todos los pares de objetos. Si, por ejemplo, todos los pares de objetos del conjunto A, B, C (es decir, $\overline{AB}$, $\overline{AC}$, $\overline{BC}$) se clasifican-ordenan, entonces todos los pares de objetos pueden también compararse. Suponga que todos los pares se clasifican como $\overline{AB} = 1$, $\overline{AC} = 2$, y $\overline{BC} = 3$ (donde 1 es el más parecido). Claramente, el par $\overline{AB}$ es más parecido que el par $\overline{AC}$, el par $\overline{AB}$ es más parecido que el par $\overline{BC}$ y el par $\overline{AC}$ es más parecido que el par $\overline{BC}$.

Se pueden utilizar diversos procedimientos para obtener percepciones de los encuestados de las similitudes entre estímulos. Cada procedimiento se basa en la noción de que las diferencias relativas entre cualquier par de estímulos deben medirse de tal forma que el investigador pueda determinar si el par es más o menos parecido que cualquier otro par. A continuación discutimos tres procedimientos habitualmente utilizados para obtener las percepciones de similitudes de los encuestados: comparación pareada de objetos, matriz de confusión y medidas derivadas.

Comparación pareada de objetos El método más ampliamente utilizado para obtener *juicios de similitud* es el método de pareado de objetos, en el cual se pregunta al encuestado simplemente que ordene o califique todos los pares de objetos. Si tenemos el estímulo A, B, C, D y E, podremos clasificar los pares $\overline{AB}$, $\overline{AC}$, $\overline{AD}$, $\overline{AE}$, $\overline{BC}$, $\overline{BD}$, $\overline{BE}$, $\overline{CD}$, $\overline{CE}$ y $\overline{DE}$ desde el más similar al menos similar. Si, por ejemplo, al par $\overline{AB}$ se le da la calificación de 1, asumiríamos que el encuestado ve ese par como contención de los dos estímulos que son más parecidos, en contraste con el resto de los otros pares. Este procedimiento ofrecería una medida no métrica de similitud. Las medidas métricas de similitud implicarían una calificación de similitud (es decir, de 1 «muy similar» a 10 «no tan similar»). Ambas formas (métrica o no métrica) pueden ser utilizadas en la mayoría de los programas de análisis multidimensional.

Matriz de confusión El emparejamiento (o «confusión») de estímulos I con estímulos J se considera indicativo de similitud. También se conocen como **conglomerados subjetivos**; el procedimiento típico de obtener estos datos es situar los objetos cuya similitud se va a medir (por ejemplo, 10 chocolatinas) en tarjetas, sea descriptivamente o mediante fotografías. Al encuestado se le pide que clasifique en montones todas las tarjetas de tal forma que todas las tarjetas de un montón representen chocolatinas similares. Algunos investigadores dicen a los encuestados que las coloquen en un número fijo de montones; otros dicen que las coloquen en tantos montones como el encuestado quiera. En cada situación, los datos resultan en una matriz de similitudes agregadas similar a una tabla de tabulación cruzada. Estos datos indican qué productos aparecen juntos más a menudo y por tanto se les considera como más similares. La recogida de datos de esta forma permite sólo el cálculo agregado de similitud, dado que las respuestas de todas las personas se combinan para obtener matrices de similitud.

Medidas derivadas Las medidas derivadas de similitud se basan habitualmente en puntuaciones dadas por los encuestados a estímulos. Por ejemplo, se pregunta a los sujetos que evalúen tres estímulos (cereza, fresa y soda lima-limón) sobre un número de atributos (dieta-no dieta, dulce-agrio, sabor fuerte-ligero) utilizando análisis diferenciales semánticos. Las respuestas se evaluarían para

cada encuestado (es decir, correlación, índice de acuerdo) para crear medidas de similitud entre los objetos. Aquí hay que hacer tres importantes supuestos:

1. El investigador ha seleccionado las dimensiones apropiadas a medir.
2. Las escalas pueden ser ponderadas (sea igualmente o desigualmente) para conseguir datos de similitud para un sujeto o grupo de sujetos
3. Incluso si se pudiera determinar la ponderación de las escalas, todos los individuos tendrían las mismas ponderaciones.

De los tres procedimientos que hemos discutido, la medida derivada es el método menos deseable en el espíritu del análisis multidimensional que pretende que la evaluación de los objetos sea hecha de tal forma que el investigador tenga una influencia mínima.

Recogida de datos de preferencia

La preferencia implica que los estímulos se juzgarían en términos de relaciones dominantes; esto es, los estímulos se ordenan en términos de la preferencia por alguna propiedad. Por ejemplo, la marca A es preferida a la marca C. Los dos procedimientos más comunes para obtener datos de preferencia son la clasificación directa y las comparaciones pareadas.

Clasificación directa Cada encuestado clasifica los objetos de más preferido a menos preferido. Se trata de un método muy popular de recogida de datos de similitud no métricos dado que es fácil de administrar para un número moderado de objetos. Es bastante parecido en concepto al procedimiento del conglomerado subjetivo ya discutido, sólo que en este caso a cada objeto se le debe dar una clasificación única.

Comparaciones pareadas A un encuestado se le presentan todos los pares posibles y se le pregunta que indique qué miembro de cada par es el preferido. De esta forma, el investigador recoge datos explícitos para cada comparación, lo que permite más detalle que las simples calificaciones directas. La principal desventaja de este método es el gran número de tareas que hay que realizar incluso con un número relativamente pequeño de objetos. Por ejemplo, 10 objetos resultan en 90 comparaciones pareadas, que son demasiadas para la mayoría de las situaciones de investigación. Nótese que las comparaciones pareadas se utilizan también en la recogida de datos de similitud, como se ilustró en el ejemplo del principio del capítulo.

Datos de preferencia frente a datos de similitud

Los datos de preferencia permiten al investigador ver la localización de los objetos en un mapa perceptual en el cual la distancia refleja las diferencias en las preferencias. Este procedimiento es útil porque la percepción de un individuo sobre un objeto en un contexto de preferencia puede ser diferente de la que se realiza en un contexto de similitud; esto es, una dimensión particular puede ser muy útil en la descripción de las diferencias entre dos objetos pero puede no ser consecuente en la determinación de una preferencia. Por tanto, dos objetos podrían ser percibidos como diferentes en un mapa de similitud, y, sin embargo, aparecer como similares en un mapa preferencial. Así, resultarían dos mapas bastante diferentes, de tal forma que dos marcas de chocolatinas distintas estarían apartadas en un mapa de similitud pero, con preferencia equivalente, estarían situadas muy cerca la una de la otra en un mapa de preferencia.

En resumen, los procedimientos de recogida de datos para datos tanto de similitud como de preferencia tienen el propósito común de obtener series de respuestas unidimensionales que representan los juicios de los encuestados. Estos juicios servirán después como datos de muchos procedimientos de análisis multidimensional que definen la pauta de multidimensionalidad subyacente en relación con estos juicios.

Tercer paso: Supuestos del análisis multidimensional

El análisis multidimensional, aunque no tiene supuestos restrictivos en la metodología, tipo de datos o forma de las relaciones entre las variables, requiere que el investigador acepte varios principios acerca de la percepción, entre ellos:

1. *Variación en la dimensionalidad*—cada encuestado no percibirá que un estímulo tenga la misma dimensionalidad (aunque se piense que la mayor parte de la gente juzga en términos de un limitado número de características o dimensiones). Por ejemplo, alguien puede evaluar un coche en términos de su potencia o su diseño, mientras que otros no consideran estos factores sino que lo evalúan en términos del coste y del bienestar interior.
2. *Variación en importancia*—los encuestados no necesitan asignar el mismo nivel de importancia a una dimensión, incluso si todos ellos perciben esta dimensión. Por ejemplo, dos encuestados perciben un refresco de cola en términos de su nivel de gas, pero alguien puede considerar esta dimensión sin importancia mientras que otros pueden considerarla muy importante.
3. *Variación en el tiempo*—los juicios de un estímulo en términos tanto de dimensiones o niveles de importancia no tienen por qué permanecer estables en el tiempo. En otras palabras, no puede esperarse que los encuestados mantengan las mismas percepciones durante largos períodos de tiempo.

A pesar de estos supuestos y de las diferencias que podemos esperar entre los individuos, el análisis multidimensional intenta representar percepciones espacialmente, de tal forma que pueda examinarse cualquier relación subyacente. El propósito de emplear una técnica de análisis multidimensional reside no sólo en entender a cada individuo por separado, sino también en identificar las percepciones compartidas y las dimensiones evaluativas dentro de una muestra de encuestados.

Cuarto paso: Obtención de soluciones de análisis multidimensional y valoración del ajuste conjunto

La variedad de programas informáticos que realizan análisis multidimensional está aumentando rápidamente, particularmente para ser utilizados por computadores personales. Hoy en día se encuentran programas de análisis multidimensional en todos los principales programas estadísticos. Nosotros empleamos una serie de aplicaciones de análisis multidimensional para ilustrar los tipos de datos de entrada, los diferentes tipos de representaciones espaciales e interpretaciones alternativas. Nuestro objetivo aquí es proporcionar una visión del análisis multidimensional que permita una rápida compresión de las diferencias entre estos programas. Sin embargo, como con otras técnicas multivariantes, hay un desarrollo continuo tanto en conocimientos como en aplicación. Por tanto, remitimos al usuario interesado en aplicaciones específicas de programas a otros textos dedicados exclusivamente al análisis multidimensional [9, 10, 16, 18, 23].

Determinación de la posición de un objeto en un mapa perceptual

La primera tarea del cuarto paso implica la situación de objetos que reflejen mejor las evaluaciones de similitud ofrecidas por los encuestados (véase Figura 10.4). Los programas del análisis multidimensional siguen un proceso común de determinación de posiciones óptimas. Este proceso puede describirse en cuatro pasos:

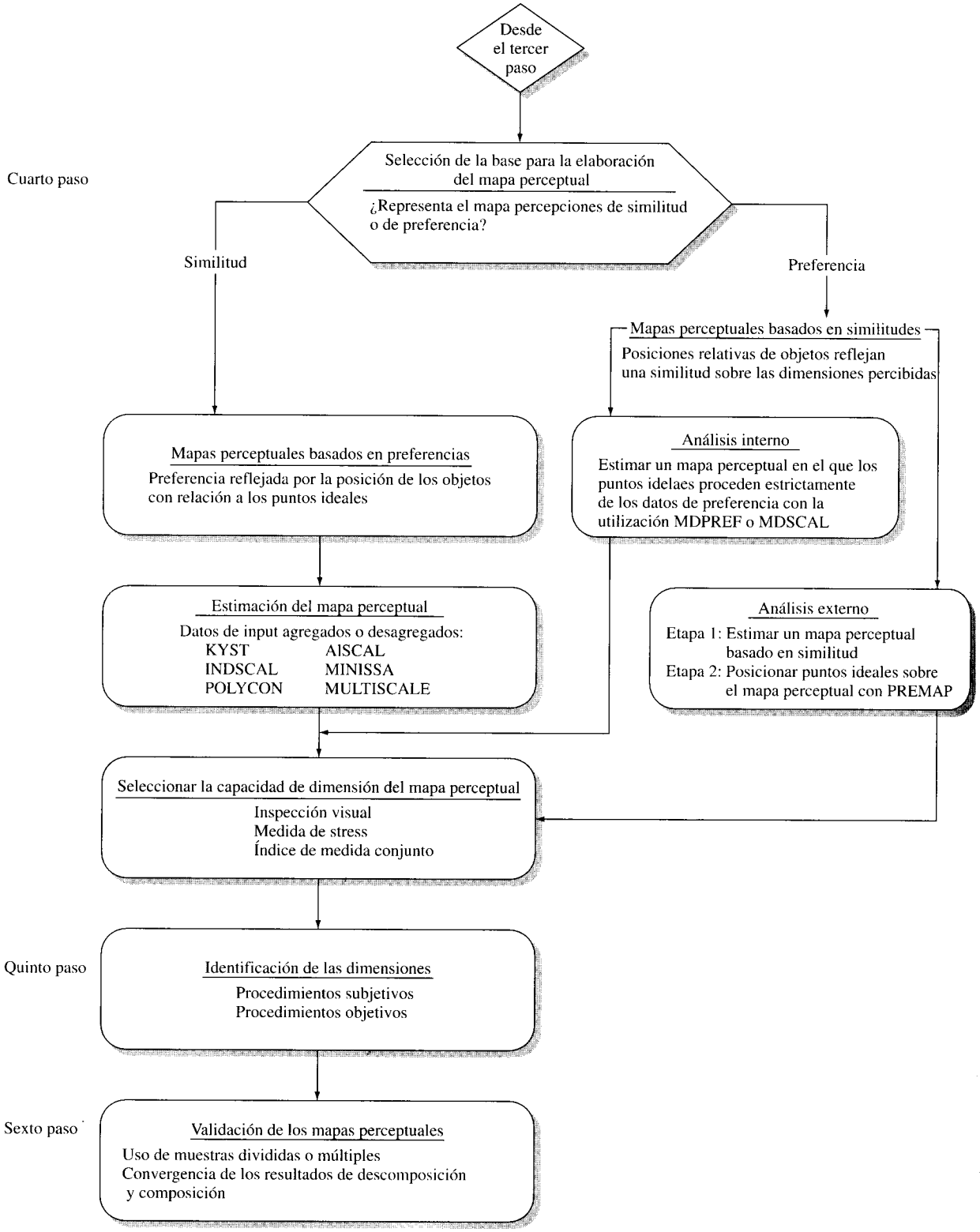


Figura 10.4. Pasos 4-6 en el diagrama de decisión del análisis multidimensional.

Paso 1: seleccionar una configuración inicial de estímulos (S_k) respecto a una **dimensionalidad inicial** (t) deseada. Existen varias opciones para obtener las configuraciones iniciales. Las dos más ampliamente utilizadas son configuraciones, bien aplicadas por el investigador basándose en datos previos, o generadas por puntos pseudoaleatorios desde una distribución multivariante aproximadamente normal.
Paso 2: calcular las distancias entre los puntos de estímulo y comparar las relaciones (observadas frente a estimadas) con una medida de ajuste. Una vez que se ha encontrado la configuración, las distancias de estímulos entre puntos (d_{ij}) en las configuraciones iniciales se comparan con las medidas de distancia ($\hat{d}_{ij}$) de juicios de similitud (s_{ij}). Las dos distancias medidas se componen entonces mediante una medida de ajuste, normalmente una medida de stress. (Las medidas de ajuste se discutirán en una sección posterior.)
Paso 3: si la medida de ajuste no llega a un valor límite seleccionado, hay que encontrar una nueva configuración para la cual la medida de ajuste se minimice aún más. El programa determina las direcciones en las cuales puede obtenerse la mejora en el ajuste y a continuación mueve los puntos de la configuración en esas direcciones mediante pequeños incrementos.
Paso 4: una vez que se ha alcanzado una medida del stress satisfactoria, la dimensionalidad se reduce a uno y el proceso se repite hasta que se alcance la menor dimensionalidad con una medida aceptable del ajuste.

La necesidad de un programa informático frente a los cálculos manuales se pone de manifiesto a medida que el número de objetos y dimensiones aumenta. Por ejemplo, si tenemos que evaluar 10 productos, cada encuestado debe clasificar los 45 pares de productos desde los más parecidos (1) a los menos parecidos (45). Colocamos los 10 puntos (representando los 10 productos) aleatoriamente en una hoja de papel milimetrado y a continuación medimos las distancias entre dos pares de puntos (45 distancias). A continuación, calculamos el stress de la solución, a medida que muestra la consonancia del orden-clasificación entre las distancias euclídeas (línea directa) de los 45 objetos dibujados y las 45 clasificaciones originales. Si las distancias en línea recta no concuerdan con los rangos originales, debemos modificar los diez puntos e intentarlo de nuevo. El proceso se hace intratable a medida que el número de objetos aumenta y las diferencias en percepción y las dimensiones utilizadas en la evaluación crecen. El computador se utiliza sólo para reemplazar los cálculos manuales y permitir una solución más detallada y precisa.

El criterio principal para encontrar la mejor representación de los datos en todos los casos es la conservación de las relaciones ordenadas entre la clasificación original de los datos y las distancias obtenidas entre los puntos. La medida de stress es simplemente una medida de lo bien (o mal) que las distancias representadas en un mapa concuerdan con las clasificaciones dadas por los encuestados.

Al evaluar un mapa perceptual, el investigador debe ser consciente de las **soluciones degeneradas**, que se derivan de los mapas perceptuales que no son representaciones precisas de las respuestas de similitud. Muy a menudo están provocadas por inconsistencias en los datos o una incapacidad del programa de análisis informático para alcanzar una solución estable. Están caracterizadas muy a menudo tanto por una pauta circular, en la cual todos los objetos se muestran igualmente similares, o una solución de conglomerado, en la cual los objetos están agrupados en dos extremos de una dimensión única. En ambos casos, el programa informático es incapaz de diferenciar entre los objetos por alguna razón. El investigador debería entonces volver a examinar el diseño de la investigación para ver dónde se ha producido la inconsistencia.

Selección de la dimensionalidad del mapa perceptual

El objetivo de este paso es la selección de una configuración espacial en un número especificado de dimensiones. La determinación de cuántas dimensiones se encuentran efectivamente representadas en los datos se investiga normalmente a través de una de estas tres aproximaciones: la evaluación subjetiva, los gráficos de caída de las medidas de stress o un índice de ajuste conjunto.

El mapa espacial es un buen punto de partida para la evaluación. El número de mapas necesarios para la interpretación depende del número de dimensiones. Se produce un mapa para cada com-

binación de dimensiones. Un objetivo del investigador sería la obtención del mejor ajuste con el menor número posible de dimensiones. La interpretación de soluciones obtenidas en más de tres dimensiones es extremadamente difícil y normalmente no merece la pena la mejora en el ajuste. Normalmente, el investigador hace una **evaluación subjetiva** de los mapas perceptuales y determina si la configuración parece razonable. Esta evaluación es importante porque en una etapa final será necesario explicar e interpretar las dimensiones.

Una segunda aproximación es utilizar una **medida de stress**, que indica la proporción de la variación de las **disparidades** no tenidas en cuenta por el análisis multidimensional. Esta medida varía de acuerdo con el tipo de programa y los datos que se van a analizar. El stress de Kruskal es la medida más utilizada en la determinación de un modelo de buen ajuste. Se define por:

$$\text{Stress} = \sqrt{\frac{(d_{ij} - \hat{d}_{ij})^2}{(d_{ij} - \bar{d})^2}}$$

donde

$\bar{d}$ = la distancia media ($\sum d_{ij}/n$) sobre el mapa
$\hat{d}_{ij}$ = distancia obtenida a partir de los datos de similitud
d_{ij} = distancias originales facilitadas por los encuestados

El valor del stress se hace más pequeño a medida que la $\hat{d}_{ij}$ se aproxima al d_{ij} original. El stress se minimiza cuando los objetos están situados en una configuración de tal forma que las distancias entre los objetos se ajusten mejor a las distancias originales.

Sin embargo, uno de los problemas que se encuentran al utilizar el stress es análogo al del R^2 en la regresión múltiple; esto es, que el stress siempre mejora con mayores dimensiones. (Recuérdese que el R^2 siempre aumenta con variables adicionales.) Se debe considerar entonces la relación entre el ajuste de la solución y el número de dimensiones. Como se hizo para la extracción de factores en el análisis factorial, podemos realizar un gráfico entre el valor del stress frente al número de dimensiones para determinar el mejor número de dimensiones para utilizarse en el análisis [18]. Por ejemplo, en el gráfico de caída en la Figura 10.5, el codo indica que existe una mejora sustancial en la bondad del ajuste cuando el número de dimensiones se aumente de 1 a 2. Por tanto, el mejor ajuste se obtiene con un número relativamente bajo de dimensiones.

A veces se utiliza un índice de correlación al cuadrado como **índice de ajuste**. Puede interpretarse como la proporción de variación de las disparidades (datos óptimamente analizados) tenidos en cuenta por el procedimiento de análisis dimensional. En otras palabras, es una medida de lo bien que se ajustan los datos al modelo de análisis multidimensional. La medida del R^2 del análisis multidimensional representa esencialmente la misma medida de variación que en otras técnicas multivariantes. Por tanto, es posible utilizar criterios de medida similares; esto es, se consideran aceptables medidas de 0,60 o mayores. Desde luego, cuanto mayor sea el R^2, mejor será el ajuste.

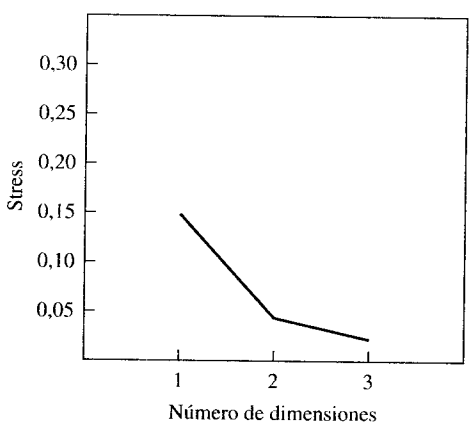


Figura 10.5. Uso de un gráfico de caída para determinar la dimensionalidad apropiada.

Incorporación de las preferencias en el análisis multidimensional

Hasta este momento, nos hemos concentrado en desarrollar mapas perceptuales basados en juicios de similitud. Sin embargo, los mapas perceptuales pueden obtenerse también a partir de las preferencias. El objetivo es determinar la combinación preferida de características para un conjunto de estímulos que predice preferencias, dado un conjunto de configuración de objetos [8, 9]. Al hacerlo, se desarrolla un espacio conjunto que representa tanto los objetos (estímulos) como los sujetos (puntos ideales). Un supuesto crítico es la homogeneidad de la percepción de los individuos para el conjunto de objetos, lo que permite que se atribuyan todas las diferencias a las preferencias, no a las diferencias perceptuales.

Puntos ideales

El término **punto ideal** a veces ha sido mal entendido o confundido. Podemos suponer que si localizamos (en el mapa perceptual estimado) el punto que representa la combinación más preferida de atributos percibidos, hemos identificado la posición de un objeto ideal. Igualmente, podemos suponer que la posición de este punto ideal (en relación con los otros productos en el mapa perceptual derivado) define preferencias relativas, de tal forma que los productos que estén lejos de este ideal serían menos preferidos. Un punto ideal se sitúa de tal forma que la distancia de cualquier objeto a este punto expresa cambios en las preferencias. Considérese, por ejemplo, la Figura 10.6. Cuando los datos de preferencia sobre las seis chocolatinas se obtuvieron de la persona indicada por el punto (·), el punto (·) estaba situado de tal forma que el aumento de la distancia con este punto indica una disminución de la preferencia. Uno puede suponer que este orden de preferencias de la persona es C, F, D, E, A, B. Suponer que la chocolatina ideal está exactamente en el punto (·) o incluso más allá (en la dirección mostrada por la línea punteada desde el origen) puede ser una equivocación. El punto ideal simplemente define la relación de preferencia ordenada por el encuestado entre el conjunto de seis chocolatinas. Aunque los puntos ideales pueden no ofrecer mucha ayuda, los conglomerados de ellos pueden ser muy útiles en la definición de segmentos. Muchos encuestados con puntos ideales en el mismo área general, representan un mercado potencial de segmentos de personas con preferencias similares, como se indica en la Figura 10.7.

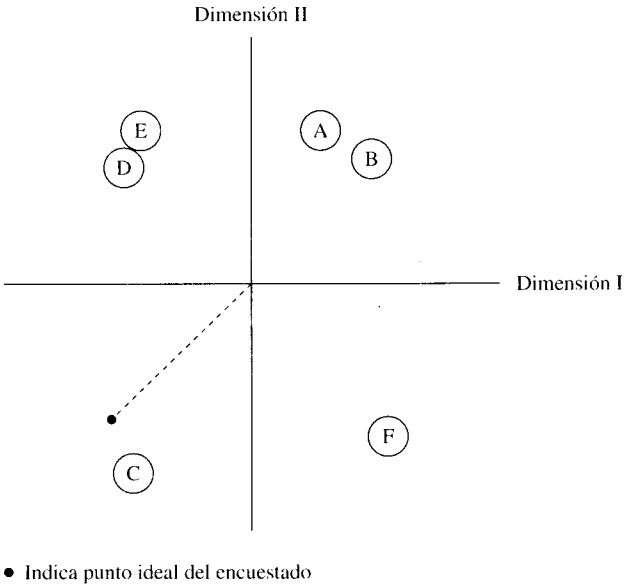


Figura 10.6. Un punto ideal de un encuestado dentro del mapa perceptual.

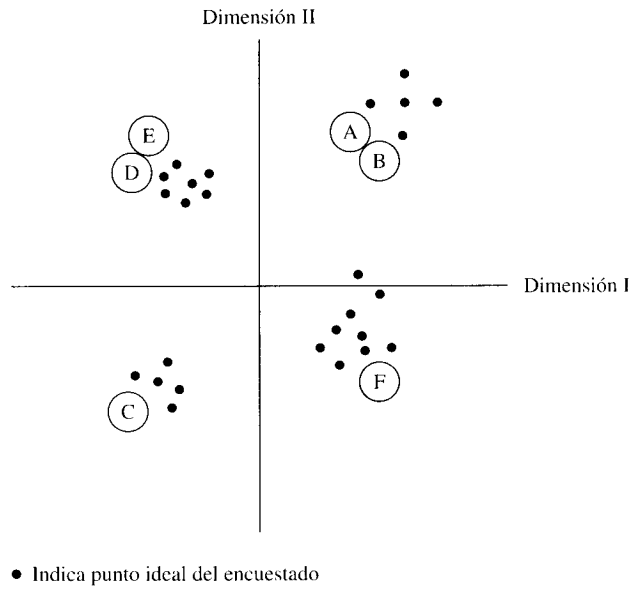


Figura 10.7. Incorporando puntos ideales múltiples en el mapa perceptual.

Generalmente, se han utilizado dos aproximaciones para determinar los puntos ideales: estimaciones implícitas y explícitas. La estimación explícita parte de las respuestas directas de los sujetos. Este procedimiento puede implicar preguntar al sujeto que califique un ideal hipotético sobre los mismos atributos a partir de los cuales se han calificado otros estímulos. Alternativamente, se pregunta al encuestado que incluya entre los estímulos utilizados para recoger los datos de similitudes, un estímulo ideal hipotético (por ejemplo, marca, imagen).

Cuando preguntamos a los encuestados que conceptualicen un ideal, normalmente nos encontraremos con problemas. A menudo, el encuestado conforma el ideal en los extremos de las calificaciones específicas utilizadas o como si fueran similares al producto más preferido de entre aquéllos con los que ha tenido alguna experiencia. También, el individuo debe pensar no en términos de similitudes, sino en términos de preferencias, lo que a menudo es difícil con objetos relativamente desconocidos. Con frecuencia, estos problemas de percepción llevan al investigador a utilizar estimaciones implícitas del punto ideal.

Existen varios procedimientos para localizar implícitamente los puntos ideales (véase la sección siguiente para una descripción más detallada). El supuesto básico subyacente bajo la mayoría de los procedimientos es que las medidas derivadas de posiciones espaciales de los puntos ideales son máximamente consistentes con las preferencias de los encuestados individuales. Srinivasan y Schocker [25] suponen que el punto ideal de todos los pares de estímulos se determina de tal forma que imcumple con menor daño la restricción de que estará más cerca de los más preferidos en cada par en lugar de ser el menos preferido.

En resumen, existen varias formas de aproximar la estimación de los puntos ideales, y no se ha demostrado la existencia de un método mejor. La elección depende de la habilidad del investigador y del procedimiento de análisis multidimensional elegido.

Situación del punto ideal

Para situar implícitamente el punto ideal a partir de los datos de preferencia, se puede desarrollar un análisis interno y externo. El análisis interno de los datos de preferencia hace referencia al desarrollo de un mapa espacial conjunto de los estímulos y puntos del sujeto (o vectores) solamente a partir de los datos de preferencia. El análisis externo de preferencia utiliza una configuración preespecificada de objetos para después intentar situar a posteriori los puntos ideales dentro de este mapa perceptual. Cada aproximación tiene ventajas e inconvenientes, que se discuten en las secciones siguientes.

Análisis interno El análisis interno debe basarse en ciertos supuestos respecto a la obtención tanto del mapa perceptual de estímulos, como de puntos ideales. Las posiciones del objeto se calculan sobre la base de datos de **exposición** de preferencias para cada individuo. Los resultados reflejan dimensiones perceptuales que se ven «estiradas» y ponderadas para predecir la preferencia. Una característica de los métodos de estimación interna es que normalmente emplean un vector de representaciones del punto ideal (véase la sección siguiente para una discusión de representaciones de vector frente a punto), mientras que los modelos externos pueden estimar representaciones tanto de vectores como de puntos.

Como ejemplo de esta aproximación, MDPREF [5] ó MDSCAL [17], dos de los programas de este tipo más utilizados, permiten al usuario encontrar configuraciones de estímulos y puntos ideales. Al hacerlo, el investigador debe suponer: (1) que no existe ninguna diferencia entre los objetos, (2) configuraciones separadas para cada objeto, o (3) una configuración única con puntos ideales individuales. Al recoger datos de preferencia, el investigador puede representar tanto estímulos como escuestados sobre un único mapa perceptual.

Análisis externo El análisis externo de los datos de preferencia se refiere a un ajuste de puntos ideales (basados en datos de preferencia) a un espacio de estímulos desarrollado a partir de datos de similitudes obtenidos de los mismos sujetos. Por ejemplo, podemos analizar datos de similitud individualmente, examinar los mapas individuales para encontrar percepciones comunes, y a continuación analizar los datos de preferencia para cualquier grupo identificado de esta forma. Si se sigue este procedimiento, el investigador tiene que recoger tanto datos de preferencia como de similitud para conseguir un análisis externo.

El PREFMAP [6] fue desarrollado exclusivamente para realizar análisis externos de datos de preferencia. Dado que la matriz de similitud define los objetos en el mapa perceptual, el investigador puede definir a continuación tanto los descriptores de atributos (suponiendo que el espacio perceptual es el mismo que las dimensiones evaluativas) como puntos ideales para los individuos. El PREFMAP ofrece estimaciones para un número de puntos ideales de diferente tipo, cada uno de ellos basado en distintos supuestos acerca de la naturaleza de las preferencias (es decir, representaciones de vector frente a punto o ponderaciones de dimensión iguales frente a diferentes).

Elegir entre el análisis interno y externo Generalmente [9, 10, 23] se prefiere el análisis externo al interno. Esta conclusión se basa en las dificultades de cálculo de los procedimientos de análisis interno y en la confusión de diferencias en preferencia con diferencias en percepción. Además, las salidas sobre las dimensiones percibidas pueden cambiar a medida que uno se desplaza desde el espacio perceptual (¿son los estímulos similares o no?) al espacio evaluativo (¿qué estímulo es el más preferido?).

Ilustramos el procedimiento de una estimación externa en nuestro ejemplo de la elaboración de un mapa perceptual con análisis multidimensional al final de este capítulo.

Representaciones vector frente a representaciones punto

La discusión de la elaboración del mapa perceptual de los datos de preferencia ha hecho hincapié en un punto ideal que representa la relación del orden de preferencias de un individuo para un conjunto de estímulos. El método de representación del punto ideal de más fácil comprensión consiste en utilizar la medida de distancia en línea recta (euclídea) de un orden de preferencia desde el punto ideal a todos los puntos que representan los objetos. Estamos suponiendo que la dirección de la distancia desde el punto ideal no es crítica, sólo representa la distancia relativa. En la Figura 10.8 se muestra un ejemplo. Aquí, el punto ideal tal y como se ha posicionado, indica que el objeto más preferido es E, seguido de C, a continuación B, D, y finalmente A.

El punto ideal también puede mostrarse como un **vector**. Para calcular las preferencias con esta aproximación, se dibujan líneas perpendiculares desde los objetos al vector. La preferencia aumenta en la dirección en que el vector está apuntando. Las preferencias pueden leerse directamente a partir del orden de las proyecciones. La Figura 10.9 ilustra el enfoque del vector para dos sujetos fren-

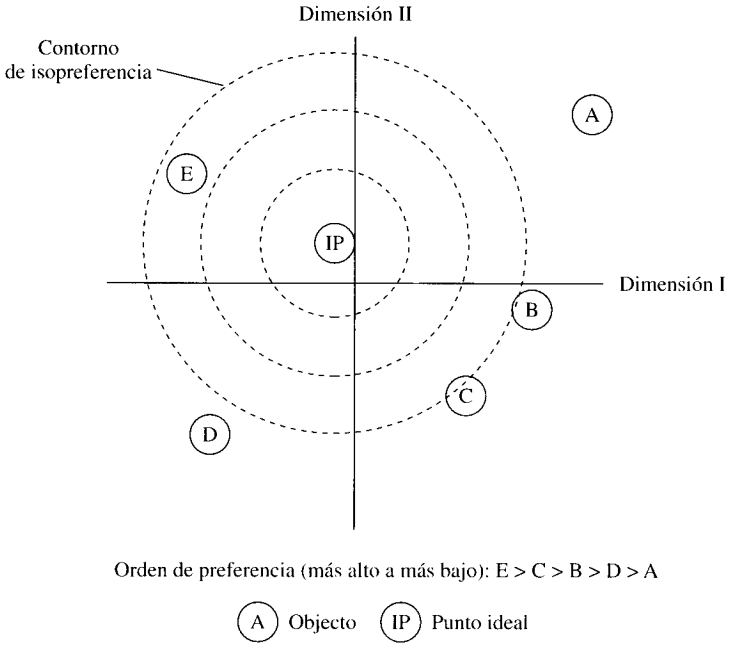


Figura 10.8. Representación de un punto ideal.

te al mismo conjunto de posiciones de estímulos. Para el sujeto 1, el vector tiene la dirección de una preferencia más baja en la esquina inferior izquierda a una preferencia más elevada en la esquina superior derecha. Cuando se realiza la proyección de cada objeto, el orden de preferencia (del mayor al menor) es A, B, C, E y D. Sin embargo, los mismos objetos tienen un orden de preferencia bastante diferente para el sujeto 2. Para el sujeto 2 el orden de preferencia va desde el más preferido, E, al menos preferido, C. De esta forma, un solo vector puede representar a cada sujeto. En el enfoque de vector, no hay un único punto ideal, pero se asume que el punto ideal está a una distancia infinita externa al vector.

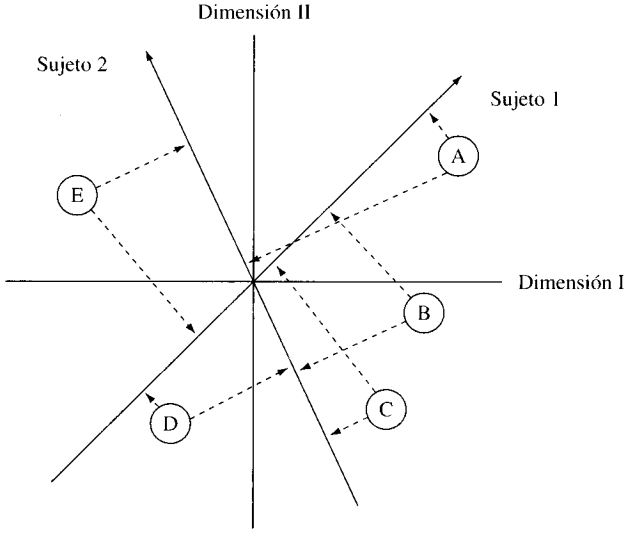
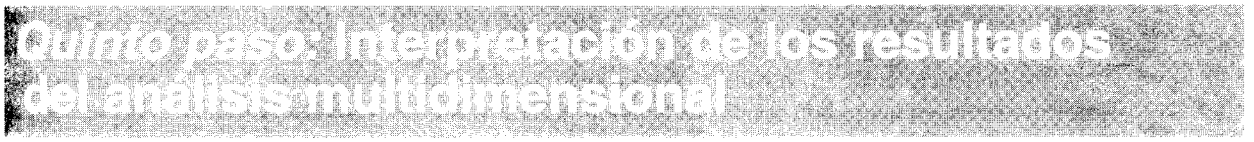


Figura 10.9. Representaciones de vector de dos puntos ideales: sujetos 1 y 2.

Aunque ni la representación de vector ni la de punto puede indicar qué combinaciones de atributos son más preferidas, estas observaciones a menudo no se ven corroboradas por una experimentación ulterior. Por ejemplo, Raymond [22] cita un ejemplo en el cual se deducía la conclusión de que la gente prefería los *brownies* debido al grado de humedad y al contenido de chocolate. Cuando los técnicos de alimentación aplicaron este resultado en el laboratorio, encontraron que sus *brownies* hechos para la especificación experimental eran de chocolate blanco. Uno no puede asumir siempre que las relaciones encontradas son independientes o lineales, o que se mantienen en el tiempo, como se apuntó previamente. Sin embargo, el análisis multidimensional es un comienzo en la comprensión de las percepciones y supone un avance considerable en nuestro conocimiento tanto de la metodología como de la percepción humana.

Resumen

Los datos de preferencia se examinan mejor utilizando el análisis externo como un medio de entender mejor tanto las diferencias perceptuales entre objetos basados en juicios de similitud y elecciones de preferencia hechas dentro de un mapa perceptual de objetos. De esta forma, el investigador puede distinguir entre ambos tipos de evaluaciones perceptuales y entender más precisamente las percepciones de los individuos.



Una vez obtenido el mapa perceptual, los dos enfoques —composición y descomposición— divergen de nuevo en su interpretación de los resultados. Para los métodos de composición, el mapa perceptual debe ser validado con otras medidas de percepción, dado que las posiciones están totalmente definidas por los atributos especificados por el investigador. Por ejemplo, los resultados del análisis discriminante se pueden aplicar a un nuevo conjunto de objetos o encuestados, evaluando la capacidad de diferenciar con estas nuevas observaciones.

Para los métodos de descomposición, el asunto más importante es la descripción de las dimensiones perceptuales y su correspondencia a los atributos. Más adelante se discutirán varias técnicas descriptivas para «etiquetar» las dimensiones, así como para integrar preferencias (para objetos y atributos) con los juicios de similitud. De nuevo, en línea con sus objetivos, los métodos de descomposición proporcionan una visión inicial de las percepciones desde la que pueden surgir perspectivas más formalizadas.

Dado que en otros capítulos del texto se han tratado muchas de las técnicas de composición, lo que queda del capítulo se centra en los métodos de descomposición, fundamentalmente las diversas técnicas utilizadas en el análisis multidimensional. Una excepción notable es la discusión de un enfoque de composición —análisis de correspondencias— que, en cierto grado, salva la brecha entre las dos aproximaciones para su flexibilidad y su método de interpretación.

Identificación de las dimensiones

Tal y como se discute en el Capítulo 3, la identificación de las dimensiones subyacentes es a menudo una tarea difícil. Las técnicas de análisis multidimensional no tienen un procedimiento específico para identificar las dimensiones. El investigador, una vez desarrollados los mapas con una dimensionalidad seleccionada, puede adoptar diversos procedimientos, tanto subjetivos como objetivos.

Procedimientos subjetivos

La interpretación siempre debe incluir algún elemento de juicio del investigador o del encuestado, y en muchos casos esto se revela adecuado para las cuestiones abiertas. Un método muy sim-

ple, aunque efectivo, es que el encuestado etiquete mediante inspección visual las dimensiones del mapa perceptual. Se puede preguntar a los encuestados que interpreten subjetivamente la dimensionalidad mediante la inspección de los mapas, o un conjunto de «expertos» puede evaluar e identificar las dimensiones. Aunque no hay un intento de vincular cuantitativamente las dimensiones a los atributos, este enfoque puede ser el mejor si se cree que las dimensiones son altamente intangibles, afectivas o emocionales, en contenido, de tal forma que no puedan deducirse descriptores adecuados.

De forma similar, los investigadores pueden describir las dimensiones en términos de características conocidas (objetivas). De esta forma, se hace la correspondencia entre las dimensiones objetiva y perceptual directamente, aunque estas relaciones no son un resultado de una interacción con el encuestado sino del juicio del investigador.

Procedimientos objetivos

Como complemento a los procedimientos subjetivos, se han desarrollado varios métodos más formalizados. El método más ampliamente utilizado, PROFIT (PROPERTY FITTING), recoge calificaciones de atributos para cada objeto y a continuación busca la mejor correspondencia de cada atributo para el espacio perceptual derivado. El intento es identificar los atributos determinantes en los juicios de similitud hechos por los individuos. Se dan las medidas de ajuste para cada atributo, así como su correspondencia con las dimensiones. El investigador puede entonces determinar qué atributos describen mejor las posiciones perceptuales y son ilustrativas de las dimensiones. La necesidad de correspondencia entre los atributos y las dimensiones definidas disminuye con el uso de resultados métricos, en la medida que las dimensiones pueden ser modificadas libremente sin cambios en la interpretación.

Elección entre procedimientos objetivos y subjetivos

Tanto para los procedimientos objetivos como para los subjetivos, se debe recordar que aunque la dimensión puede representar un atributo único, habitualmente no lo hace. Un procedimiento común es recoger los datos de varios atributos, asociándolos bien subjetivamente bien empíricamente con las dimensiones donde se aplique, y determinar las etiquetas para cada dimensión utilizando múltiples atributos, de la misma forma que el análisis factorial. Muchos investigadores sugieren que la mejor alternativa consiste en utilizar los datos de atributos para ayudar a etiquetar las dimensiones. El problema, sin embargo, es que el investigador puede no incluir todos los atributos importantes en el estudio. Por tanto, puede no estar nunca totalmente seguro de que las etiquetas representen todos los atributos relevantes.

Tanto los procedimientos subjetivos como los objetivos, ilustran la dificultad de etiquetar los ejes. Esta tarea no debe dejarse para el final, en la medida en que las denominaciones de las dimensiones son esenciales para una ulterior interpretación y uso de los resultados. El investigador debe seleccionar el tipo de procedimiento que mejor se ajusta tanto a los objetivos de la investigación, como a la información disponible. Por tanto, debe plantearse tanto la obtención de denominaciones para las dimensiones, como la estimación del mapa perceptual.

Sexto paso: Validación de los resultados del análisis multidimensional

La validación en el análisis multidimensional es tan importante como en cualquier otra técnica multidimensional. Debido a la naturaleza altamente inferencial del análisis multidimensional, este esfuerzo debería dirigirse a asegurar la generalidad de resultados tanto en los objetos como en la población. Pero los esfuerzos de validación son problemáticos. El único resultado del análisis mul-

tidimensional que puede ser utilizado a efectos comparativos concierne a las posiciones relativas de los objetos. Por tanto, aunque las posiciones puedan compararse, las dimensiones subyacentes no tienen fundamento para la comparación. Si las posiciones varían, el investigador no puede determinar si los objetos se ven de forma diferente, si las dimensiones perceptuales varían o ambas cosas. Más aún, no se han incorporado e integrado en los programas informáticos métodos sistemáticos de comparación. Se deja a la improvisación del investigador, con métodos que pueden servir como líneas generales, pero que no son específicos de resultados de análisis multidimensional.

¿Qué opciones quedan disponibles? La aproximación más directa es un split —o comparación de varias muestras, en la que o bien se divide la muestra original o se recoge una nueva muestra. En cada caso, el investigador debe encontrar los medios para comparar los resultados. A menudo, la comparación entre resultados se hace visualmente o con una correlación simple de coordenadas. Hay algunos programas informáticos, como el FMATCH [24], pero el investigador debe determinar a continuación cuántas de las disparidades se deben a diferencias en las percepciones del objeto, dimensiones diferentes, o ambas.

Otro método es obtener una convergencia de los resultados del análisis multidimensional aplicando tanto métodos de descomposición como de composición a la misma muestra. Podría aplicarse en primer lugar el método de descomposición, junto con la interpretación de las dimensiones para identificar los atributos claves. A continuación, podrían aplicarse uno o más métodos de composición, en particular los del análisis de correspondencias, para confirmar los resultados. El investigador debe darse cuenta de que ésta no es una verdadera validación de los resultados como generalizables, pero confirma la interpretación de la dimensión. Desde este punto, podrían considerarse los esfuerzos de validación con otras muestras y otros objetos para demostrar la generalidad de otras muestras.

Análisis de correspondencias

Hasta este punto, hemos discutido las técnicas de descomposición tradicionales del análisis multidimensional, ¿pero qué hay que decir acerca de las técnicas de composición? En el pasado, los enfoques de composición habían descansado sobre las técnicas multivariantes tradicionales, tales como el análisis discriminante o el análisis factorial. Pero desarrollos recientes han combinado aspectos de ambos métodos y del análisis multidimensional para configurar nuevos instrumentos de elaboración de mapas perceptuales.

El **análisis de correspondencias** (CA) es una técnica de interdependencia que se ha ido haciendo más popular para la reducción dimensional y la elaboración de mapas perceptuales [1, 2, 11, 13, 19]. Es una técnica de composición debido a que el mapa perceptual se basa en la asociación entre objetos y un conjunto de características descriptivas o atributos especificados por el investigador. Entre las técnicas de composición, el análisis factorial es el más parecido, pero el análisis de correspondencia va más allá del análisis factorial. Su aplicación más directa es la representación de la «correspondencia» de categorías de variables, particularmente aquellas medidas en escalas de medida nominales. Esta correspondencia es la base del desarrollo de los mapas perceptuales. Los beneficios de CA se basan en sus capacidades únicas para representar filas y columnas, por ejemplo, etiquetas *y* atributos, en un mismo espacio.

Examinamos a continuación una situación simple de CA para lograr cierta perspectiva de sus principios básicos. A continuación, discutiremos cada uno de los seis pasos en el proceso de toma de decisiones que se introdujo en el Capítulo 1. Se pondrá un mayor énfasis en aquellos elementos únicos del CA en comparación con los métodos de descomposición del análisis multidimensional ya discutidos.

Un ejemplo sencillo de CA

Examinemos una situación sencilla como introducción al CA. En su forma más básica, CA examina las relaciones entre categorías de datos nominales en una **tabla de contingencia**, la tabu-

lación cruzada de dos variables categóricas. Por ejemplo, supongamos que las cifras de ventas de los productos A, B y C se dividen en tres categorías de edades (adultos jóvenes, con edades comprendidas entre los 18 y los 35 años; adultos con edades comprendidas entre los 36 y los 55 años; y adultos mayores de 56 años). Los datos de tabulación cruzada se muestran en la Tabla 10.2.

Los datos muestran que las unidades vendidas varían sustancialmente por productos (el producto C tiene las mayores ventas totales, el producto B las más bajas) y por grupos de edad (los adultos de edades medias compran más unidades, los jóvenes los que menos). Pero queremos identificar cualquier patrón de ventas de tal forma que podamos afirmar que los jóvenes compran más del producto X o que los adultos compran más del producto Z. Para hacer esto, necesitamos una medida de estandarización de unidades de ventas que considere simultáneamente las diferencias en las ventas para una combinación de producto específico-categoría de edad. A continuación, si todavía vemos que un cierto grupo de edad compra más unidades de un producto que las esperadas, podemos asociar ese grupo de edad con ese producto. En una representación gráfica, los grupos de edad se localizarían más cerca de productos con los cuales están altamente asociados y más lejos de grupos con asociaciones más bajas. Del mismo modo, queremos ser capaces de ver cualquier producto y ver sus asociaciones con varios grupos de edad.

Cálculo de una medida de asociación

El análisis de correspondencias utiliza uno de los conceptos estadísticos más básicos, la chi-cuadrado, para estandarizar las ventas (valores de frecuencia) y formar las bases de las asociaciones. La **chi-cuadrado** es una medida estandarizada de las frecuencias observadas de cada celda con las frecuencias esperadas de celdas. En nuestros datos de tabulación cruzada, cada celda contiene las ventas para una combinación de un producto-grupo de edad. El procedimiento de la chi-cuadrado procede a continuación en tres pasos para calcular un valor de chi-cuadrado para cada celda:

1. *Cálculo de las ventas esperadas.* El primer paso es calcular las ventas esperadas para una celda como si no existiese asociación. Las ventas esperadas se definen como la probabilidad conjunta de la combinación de columna (producto) y fila (grupo de edad). Ésta se calcula como la probabilidad marginal para el producto (ventas de ese producto para todos los grupos de edad/ventas totales para todos los grupos de edad y productos) por la probabilidad marginal para el grupo de edad (ventas de ese grupo de edad para todos los productos/ventas totales para todos los grupos de edad y productos). A continuación, este valor se multiplica por las ventas totales para todos los grupos de edad y todos los productos. Puede simplificarse cancelando términos de tal forma que la ecuación es:

$$\text{Ventas esperadas} = \frac{\text{ventas totales por categoría de edad} \times \text{ventas totales del tipo de producto}}{\text{conjunto de las ventas totales}}$$

Tabla 10.2. Datos de tabulación cruzada que detallan ventas por producto y por categoría de edad

	Ventas del producto			
Categoría de edad	*A*	*B*	*C*	*Total*
Jovenes (18 – 35 años)	20	20	20	60
Adultos (36 – 55 años)	40	10	40	90
Tercera edad (56 + años)	20	10	40	70
Total	80	40	100	220

En nuestro ejemplo, las ventas esperadas de los jóvenes que compran el producto A son de 21,28 unidades, como se muestra en el siguiente cálculo:

$$\textit{Ventas esperadas}_{\textit{jóvenes, producto A}} = \frac{60 \times 80}{220} = 21{,}82$$

Este cálculo se realiza para cada celda, con los resultados que se muestran en la Tabla 10.3.

2. *Diferencia entre los valores esperados y observados.* El siguiente paso es calcular la diferencia entre las ventas esperadas y observadas como sigue: diferencia = ventas esperadas – ventas observadas. De nuevo, en nuestro ejemplo de la celda para los jóvenes que compran el producto A, la diferencia es 1,82 (21,82 – 20,00). Una gran diferencia positiva significaría que la combinación producto-grupo de edad tiene menos ventas de las que

Tabla 10.3. Calculando los valores de similitud de la chi-cuadrado para datos de tabulación cruzada

Categoría de edad	*Ventas de producto*			
	A	*B*	*C*	*Total*
Jóvenes				
Ventas	20	20	20	60
Porcentaje de columna	25%	50%	20%	27%
Porcentaje de fila	33%	33%	33%	100%
Ventas esperadas[a]	21,82	10,91	27,27	60
Diferencia[b]	1,82	−9,09	7,27	—
Valor chi-cuadrado[c]	0,15	7,58	1,94	9,67
Similitud	−0,15	7,58	−1,94	
Adultos				
Ventas	40	10	40	90
Porcentaje de columna	50%	25%	40%	41%
Porcentaje de fila	44%	11%	44%	100%
Ventas esperadas	32,73	16,36	40,91	90
Diferencia	−7,27	6,36	0,91	—
Valor chi-cuadrado	1,62	2,47	0,02	4,11
Similitud	1,62	−2,47	−0,02	
Tercera edad				
Ventas	20	10	40	70
Porcentaje de columna	25%	25%	40%	32%
Porcentaje de fila	29%	14%	57%	100%
Ventas esperadas	25,45	12,73	31,82	70
Diferencia	5,45	2,73	−8,18	—
Valor chi-cuadrado	1,17	0,58	2,10	3,85
Similitud	−1,17	−0,58	2,10	
Total				
Ventas	80	40	100	220
Porcentaje de columna	100%	100%	100%	100%
Porcentaje de fila	36%	18%	46%	100%
Ventas esperadas	80	40	100	220
Diferencia	—	—	—	—
Valor chi-cuadrado	2,94	10,63	4,06	17,63

[a]Ventas esperadas = (Total de fila × Total de columna) ÷ Total general

Ejemplo: $\text{Celda}_{\text{jóvenes, producto A}} = (60 \times 80) \div 220 = 21{,}82$

[b]Diferencia = Ventas esperadas − Ventas reales

Ejemplo: $\text{Celda}_{\text{jóvenes, producto A}} = 21{,}82 - 20{,}00 = 1{,}82$

[c]Valor chi-cuadrado $= \dfrac{(\text{Diferencia})^2}{\text{Ventas esperadas}}$

Ejemplo: $\text{Celda}_{\text{jóvenes, producto A}} = (1{,}82)^2 \div 21{,}82 = {,}15$

cabría esperar (una asociación negativa) y una gran diferencia negativa indicaría asociaciones positivas (la celda de las ventas observadas mayor que la de ventas esperadas). Las diferencias para cada celda se muestran también en la Tabla 10.3.

3. *Calcular el valor de la chi-cuadrado.* El paso final es estandarizar las diferencias entre las celdas de tal forma que se puedan realizar fácilmente las comparaciones. La estandarización es necesaria porque sería mucho más fácil que se presentasen diferencias si las frecuencias de celdas (ventas) fueran mucho más altas comparadas con una celda con sólo unas pocas ventas. Así que, estandarizamos las diferencias para formar un valor de una chi-cuadrado dividiendo cada diferencia al cuadrado por el valor esperado de las ventas. Por tanto, el valor de la chi-cuadrado para una celda se calcula de la siguiente forma:

$$\text{Valor chi-cuadrado } (\chi^2) \text{ de una celda} = \frac{(\text{Diferencia})^2}{\text{ventas esperadas}}$$

Para nuestro ejemplo, el valor de la chi-cuadrado sería:

$$\text{Chi cuadrado } (\chi^2)_{\text{jóvenes, producto A}} = \frac{(1{,}82)^2}{21{,}82} = 0{,}15$$

Los valores de la chi-cuadrado pueden convertirse en medidas de similitud aplicando el signo opuesto de la diferencia. Por tanto, para nuestro ejemplo anterior de la celda, el valor chi-cuadrado de 0,15 podría decirse que es un valor de similitud de –0,15, dado que la diferencia era positiva. Esto es necesario porque el cálculo de la chi-cuadrado eleva al cuadrado las diferencias y los signos negativos, de tal forma que se eliminen. El resultado es una medida que actúa justamente como las medidas de similitud utilizadas en los ejemplos anteriores. Los valores negativos indican menos asociación (similitud) y los valores positivos indican una mayor asociación. Los valores de la chi-cuadrado para cada celda se muestran en la Tabla 10.3.

Las celdas con grandes valores de similitud positivos (indicando una asociación positiva) son los jóvenes-producto B (+7,58), adultos-producto A (+1,62) y tercera edad-producto C (+2,10). Cada uno de estos pares de categorías estaría muy cercano en un mapa perceptual. Las celdas con grandes valores de similitud (que significan que las ventas esperadas exceden a las reales, o una asociación negativa) serían los jóvenes-producto C (–1,94), adultos-producto B (–2,47) y tercera edad-producto A (–1,17). Donde sea posible, estas categorías deberían estar alejadas en el mapa.

Creación del mapa perceptual

Los valores de similitud (denominados chi-cuadrado) ofrecen una medida estandarizada de asociación, más parecidos a los juicios de similitud del ejemplo anterior de las chocolatinas. Con estas medidas de asociación, CA crea una medida de distancia métrica y crea dimensiones ortogonales sobre las cuales se pueden colocar las categorías para tener más en cuenta la fortaleza de la asociación representada por las distancias de la chi-cuadrado. Como hicimos en el ejemplo del análisis multidimensional, podemos considerar en primer lugar una solución unidimensional, a continuación aumentar a dos dimensiones y seguir hasta que alcancemos el máximo número de dimensiones. El máximo es uno menos el menor número de las filas o columnas. En este ejemplo, sólo podemos tener dos dimensiones (número de filas menos número de columnas menos uno = 3 – 1 = 2). El mapa perceptual es bidimensional y se muestra en la Figura 10.10.

De acuerdo con nuestro examen de las distancias chi-cuadrado, el grupo de edad de los adultos jóvenes está más cerca del producto B, los de edad mediana están más cerca del producto A y los de edad más avanzada están más cerca del producto C. Asimismo, se representan también las asociaciones negativas en las posiciones de los productos y grupos de edad. El investigador puede examinar el mapa perceptual para entender las preferencias del producto entre los grupos de edad basándose en sus patrones de ventas. Pero, como ocurre con el análisis multidimensional, no conocemos por qué existen los patrones de ventas, sino sólo cómo identificarlos.

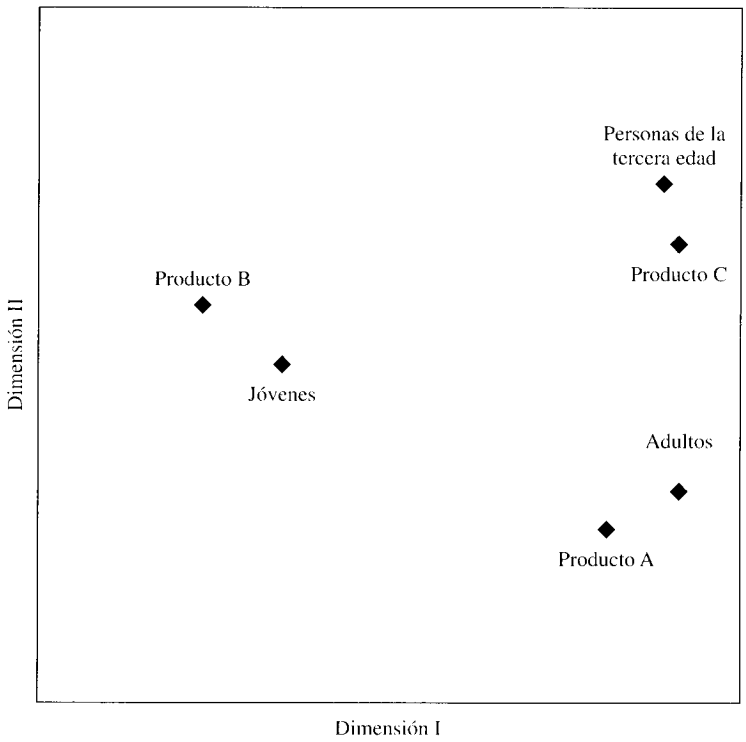


Figura 10.10. Mapa perceptual del análisis de correspondencias.

Primer paso: Objetivos del análisis de correspondencias

Los investigadores se encuentran constantemente con la necesidad de «cuantificar datos cualitativos» hallados en variables nominales. El CA difiere de otras técnicas de interdependencia en su capacidad para utilizar tanto datos no métricos como relaciones no lineales. Realiza la reducción dimensional de la misma forma que el análisis multidimensional o el análisis factorial y un tipo de elaboración de mapas perceptuales, en los que las categorías se encuentran representadas en el espacio multidimensional. La proximidad indica el nivel de asociación entre las categorías de filas o columnas. El CA puede tener dos objetivos básicos:

1. *Asociación entre categorías de columna o fila.* El CA puede utilizarse para examinar la asociación entre las categorías de sólo una fila o sólo una columna. Un uso típico es el examen de las categorías de una escala, como la escala de Likert (cinco categorías que van del «totalmente de acuerdo» al «totalmente en desacuerdo») u otras escalas cualitativas (es decir, excelente, bueno, regular, malo). Las categorías pueden compararse para ver si dos de ellas pueden ser combinadas (es decir, están muy próximas en el mapa) o si ofrecen discriminación (es decir, están localizadas separadamente en el mapa perceptual).
2. *Asociación entre categorías de filas y columnas.* En esta aplicación, el interés está en representar la asociación entre categorías de filas y columnas, de la misma forma que nuestro ejemplo de producto por grupo de edad. Este uso es más parecido a los ejemplos previos de análisis multidimensional y ha impulsado un amplio uso del CA en muchas áreas de investigación.

El investigador debe determinar los objetivos específicos del análisis debido a que ciertas decisiones se basan en el tipo de objetivos elegido. El CA ofrece una representación multivariante de interdependencia para datos no métricos que no es posible realizar con otros métodos. Pero como método de composición, el investigador debe asegurarse de que todas las variables relevantes apropiadas para

la cuestión a investigar han sido incluidas, en contraste con los métodos de análisis multidimensional de descomposición descritos anteriormente, que sólo requieren una medida conjunta de similitud.

Segundo paso: Diseño de la investigación mediante análisis de correspondencias

El análisis de correspondencias sólo requiere una matriz de datos rectangular (tabulación cruzada) de entradas no negativas. Las filas y las columnas no tienen significados predefinidos (es decir, los atributos no siempre tienen que ser filas) pero en su lugar representan las respuestas de una o más variables categóricas. Las categorías para una fila o una columna no tienen que ser necesariamente una única variable sino que pueden representar cualquier conjunto de relaciones. Como ejemplo principal tenemos el método de "escoja cualquiera"* [14, 15], en el que se da a los encuestados un conjunto de objetos y características. Los encuestados indican a continuación qué objetos, si es que hay alguno, están descritos por esas características. Nótese que el encuestado puede elegir cualquier número de objetos para cada característica, en lugar de un número ya especificado (es decir, elegir sólo el objeto que mejor se describe o el mejor de dos objetos). En esta situación, la tabla de tabulación cruzada sería el número total de veces que cada objeto está descrito por cada característica.

La tabulación cruzada de más de dos variables bajo la forma de una matriz multientrada se conoce como el **análisis de correspondencias múltiple**. En un procedimiento bastante similar al análisis de dos entradas, las variables adicionales son «ajustadas», de forma que todas las categorías están situadas en el mismo espacio multidimensional.

Tercer paso: Supuestos del análisis de correspondencias

El análisis de correspondencias comparte con las técnicas de análisis multidimensional más tradicionales una relativa libertad respecto a sus supuestos básicos. El uso de datos estrictamente no métricos en su forma más simple (datos de tabulación cruzada) representa igualmente bien relaciones lineales y no lineales. La falta de supuestos, sin embargo, no debe llevar al investigador a no esforzarse en conseguir la comparabilidad de los objetos y, dado que es una técnica de composición, a considerar la generalidad de los atributos utilizados.

Cuarto paso: Obtención de resultados con análisis de correspondencias y valoración del ajuste conjunto

Con una tabulación cruzada, las frecuencias para cualquier combinación de filas y columnas de las categorías están relacionadas con otras combinaciones basadas en frecuencias marginales. Este procedimiento proporciona una expectativa condicionada (un valor chi-cuadrado). Una vez obtenido, estos valores de la chi-cuadrado se estandarizan y se convierten en una distancia métrica, y a continuación, en un proceso mucho más parecido al análisis multidimensional, se definen soluciones de dimensiones reducidas. Estos «factores» relacionan simultáneamente filas y columnas en un único gráfico conjunto. El resultado es una representación de categorías de filas y/o columnas (es decir, marcas y atributos) en el mismo gráfico. Existen varios programas informáticos para realizar el análisis de correspondencias. Entre los programas más populares están ANACOR y HOMALS, que se encuentran en el SSPS; CA de BMDP; CORRAN y CORRESP de PC-MDS [24]; y MAPWISE [21].

Para evaluar el ajuste conjunto, el investigador debe identificar en primer lugar el número apropiado de dimensiones y su importancia. El número máximo de dimensiones que pueden ser estimadas es uno menos el número más pequeño de filas o columnas. Por ejemplo, con seis columnas y ocho filas, el número máximo de dimensiones sería cinco, dado que es seis (número de colum-

* *Nota:* Traducción de "pick any".

nas) menos uno. Los autovalores, también conocidos como valores singulares, se obtienen para cada dimensión e indican la contribución relativa de cada dimensión en la explicación de la variación en las categorías. Algunos programas, como el SPSS, introducen una medida denominada «inercia», que también mide la variación explicada y está directamente desarrollada con el autovalor. El investigador selecciona el número de dimensiones basándose en el nivel conjunto de explicación deseada de la variación y el aumento de explicación ganado por la adición de otra dimensión. Se puede decir como norma general, que aquellas dimensiones con inercia (autovalores) mayor que 0,2 deberían ser incluidas en el análisis. Como se discutió en relación con la elaboración de mapas perceptuales, utilizar una representación tridimensional o menor facilita la interpretación.

Quinto paso: Interpretación de los resultados

Una vez que se ha establecido la dimensionalidad, el investigador puede identificar una asociación de categorías con otras categorías por su proximidad después de hacer la normalización apropiada. El investigador debe seleccionar el tipo de normalización, y determinar si las comparaciones se van a hacer entre categorías de filas, categorías de columnas o categorías de filas y columnas. En muchos casos, se desean comparar categorías de filas y de columnas. Pueden existir muchos casos, sin embargo, en los cuales el interés se centra sólo en filas o columnas, como cuando se examinan las categorías de una escala para ver si pueden ser combinadas. Hasta este momento, hay un debate acerca de si es apropiado comparar entre categorías de filas y columnas. Algunos programas informáticos ofrecen un procedimiento de normalización que permite esta comparación directa. Si sólo se dispone de una fila o columna normalizada, se proponen procedimientos alternativos para hacer todas las categorías comparables [2, 21], pero todavía hay desacuerdo en su éxito [12]. En los casos para los cuales no se pueden hacer comparaciones directas, la correspondencia general todavía se mantiene y pueden distinguirse patrones específicos.

Si el investigador está interesado en definir el carácter de una o más dimensiones en términos de las categorías de filas y columnas, existen medidas descriptivas que indican la asociación de cada categoría con una dimensión específica. De carácter similar a las cargas de los factores, estas medidas detallan la medida de la asociación individualmente para cada dimensión así como colectivamente. De las medidas colectivas, se puede hacer una evaluación del ajuste para cada categoría.

Sexto paso: Validación de los resultados

La naturaleza composicional del análisis de correspondencias ofrece al investigador más posibilidades de validar los resultados. Como con todas las técnicas del análisis multidimensional, hay que hacer más énfasis en asegurar la generalización mediante análisis split o multimuestra. Sin embargo, como con otras técnicas de elaboración de mapas perceptuales, debe establecerse la generalidad de los objetos (individualmente y como un conjunto). Puede evaluarse la sensibilidad de los resultados a la adición o sustracción de un objeto, así como a la adición o la sustracción de un atributo. El objetivo es evaluar si el análisis es dependiente de sólo unos pocos objetos y/o atributos. En cada caso, el investigador debe entender el «verdadero» significado de los resultados en términos de los objetos y los atributos.

Visión de conjunto del análisis de correspondencias

El análisis de correspondencias ofrece al investigador varias ventajas. En primer lugar, puede presentarse en un espacio perceptual la tabulación cruzada simple de variables categóricas múltiples, tales como atributos de los productos versus marcas. Este enfoque permite al investigador analizar las respuestas existentes o conseguir respuestas con el tipo de medida menos restrictivo, al nivel nominal o categórico. Por ejemplo, el encuestado sólo necesita decir sí o no para un número de objetos sobre un número de atributos. A continuación, estas respuestas se pueden incluir en una ta-

bla de tabulación cruzada y ser analizadas. Otras técnicas, como el análisis factorial, requieren clasificaciones de intervalos de cada atributo para cada objeto.

En segundo lugar, CA representa no sólo las relaciones entre filas y columnas, sino también las relaciones entre las categorías tanto de las filas como de las columnas. Por ejemplo, si las columnas fueran atributos, muchos atributos cercanos tendrían todos perfiles similares para todos los productos. Con ello se forma un grupo de atributos bastante similar al factor obtenido mediante análisis de componentes principales.

Finalmente, y lo más importante, CA puede ofrecer una representación conjunta de categorías de filas y columnas en la misma dimensionalidad. Algunas modificaciones de los programas permiten comparaciones entre los puntos en las cuales la proximidad relativa está directamente relacionada con una mayor asociación entre puntos separados [1, 21]. Cuando son posibles estas comparaciones, permiten examinar las categorías de filas y columnas simultáneamente. Un análisis de este tipo permitiría al investigador identificar grupos de productos caracterizados por atributos muy relacionados.

Sin embargo, el CA también tiene desventajas o limitaciones. La primera es que la técnica es descriptiva y no del todo apropiada para la contrastación de hipótesis. Si se desea la relación cuantitativa de categorías, se sugieren métodos tales como los modelos no lineales. El CA es más apropiado para análisis exploratorio de los datos. En segundo lugar, el CA, como ocurre con muchos métodos de reducción de la dimensionalidad, no cuenta con un método para determinar concluyentemente el número de dimensiones apropiado. Como ocurre con otros métodos similares, el investigador debe sopesar la interpretabilidad con la parsimonia de la representación de los datos. Finalmente, la técnica es muy sensible a los casos atípicos, en términos tanto de filas como de columnas (es decir, atributos o marcas). También, a efectos de generalización, se pueden presentar problemas si se omite algún objeto o atributo relevante.

Ilustración del análisis multidimensional y del análisis de correspondencias

Para demostrar el uso de las técnicas de análisis multidimensional, examinaremos los datos recogidos en una serie de entrevistas con compañías representativas de una sección cruzada de clientes potenciales. En el curso del análisis de la elaboración de un mapa perceptual, aplicaremos tanto los métodos de composición como los de descomposición. La exposición examina en primer lugar los tres niveles iniciales del proceso de elaboración del modelo que son comunes a los tres métodos. La discusión se centra a continuación en los dos pasos siguientes de los métodos de descomposición. Posteriormente seguiremos con una discusión de un método de composición (CA), aplicado a la misma muestra de encuestados. El sexto paso está dedicado a la validación del análisis a través de la comparación de los resultados de ambos tipos de métodos. Finalmente, se presenta un repaso de los métodos del análisis multidimensional.

Primer paso: Objetivos del mapa perceptual

El propósito de la investigación es explorar la imagen y la competitividad de HATCO. Esta exploración incluye la identificación de las percepciones que HATCO y sus nueve competidores principales tienen en el mercado, así como la investigación de las preferencias entre los posibles clientes. Los datos se analizan en un plan de dos fases: (1) identificación de la posición de HATCO en un mapa perceptual de sus principales competidores en el mercado con un estudio de las comparaciones de las dimensiones utilizadas por los potenciales clientes, y (2) evaluación de las preferencias hacia HATCO en relación a sus principales competidores. Antes de proceder a una discusión de los resultados, describiremos brevemente el proceso de recogida de datos.

Segundo paso: Diseño de la investigación mediante el estudio del mapa perceptual

El estudio de la imagen de HATCO comprende de entrevistas en profundidad con 18 jefes de personal de nivel medio, procedentes de diferentes firmas, seleccionadas como representativas de la base de clientes potenciales existentes en el mercado. Los nueve competidores, más HATCO, representan todas las empresas más importantes de la industria y colectivamente poseen más del 85 por ciento del total de las ventas. En el curso de las entrevistas, se recogieron tres tipos de datos: juicios de similitud, calificación de atributos entre empresas y preferencias de cada empresa ante diferentes situaciones de compra.

Datos de similitud

El punto de partida de la recogida de datos fue la obtención de percepciones de los encuestados correspondientes a la similitud o disparidad entre HATCO y las nueve empresas competidoras en el mercado. Los juicios de similitud se hicieron con el enfoque de comparaciones pareadas entre objetos. Se presentaron a los encuestados 45 pares de empresas [$(10 \times 9)/2$], e indicaron la similitud de cada par de ellas sobre una escala de nueve puntos con 1 siendo «completamente diferente» y 9 siendo «muy parecido». Nótese que los valores tienen que ser transformados debido a que valores crecientes para calificaciones de similitud indican mayor similitud, lo opuesto de una medida de distancia.

Calificación de datos

Además de los juicios de similitud, se obtuvieron las calificaciones de cada empresa para ocho atributos por dos métodos. Los atributos incluían calidad del producto, orientación de la gestión, calidad del servicio, velocidad de entrega, nivel de precio, imagen de la fuerza de ventas, flexibilidad de precios e imagen de producción. En el primer método, cada empresa fue clasificada en una escala de seis puntos para cada atributo. En el segundo método, se pidió a cada encuestado que escogiese las empresas que mejor caracterizan cada atributo. Igual que en el método de «escoja una» [14, 15], el encuestado podría elegir cualquier número de empresas para cada atributo.

Evaluaciones de preferencia

Los datos finales sirven para evaluar las preferencias de cada encuestado respecto a las 10 empresas en tres situaciones de compra diferentes: una venta igual a la anterior, una venta modificada respecto a una anterior y una nueva situación de venta. En cada situación, los encuestados clasifican las empresas en orden de preferencia para ese tipo particular de compra. Por ejemplo, en las situaciones de venta igual a la anterior, el encuestado indicaba la empresa más preferida mediante la simple reordenación de los productos (orden de clasificación = 1), la siguiente más preferida (orden de clasificación = 2), y así sucesivamente. Asimismo, se recogieron preferencias similares para las situaciones de compra restantes.

Tercer paso: Supuestos básicos en la elaboración de un mapa perceptual

Los supuestos del análisis multidimensional y del análisis de correspondencias se centran principalmente en la comparabilidad y representatividad de los objetos que están siendo evaluados y de los encuestados. En relación a la muestra, el plan muestral resaltaba la importancia de obtener una muestra representativa de los clientes de HATCO. Es más, se tomó la precaución de reclutar en-

cuestados de una posición comparable y con conocimiento del mercado. Dado que HATCO y el resto de las empresas sirven a mercados claramente distintos, todas las empresas evaluadas en la elaboración del mapa perceptual deberían ser conocidas, asegurando que las posiciones discrepantes puedan atribuirse a diferencias perceptuales entre los encuestados.

Análisis multidimensional: Pasos cuarto y quinto

Una vez que se han especificado las empresas que se van a incluir en el estudio de imagen, la directiva de HATCO decidió que se iban a emplear tanto el análisis de descomposición como el análisis de composición en la construcción de los mapas perceptuales. Discutiremos en primer lugar una serie de técnicas de descomposición, y a continuación examinaremos una aproximación de composición a la elaboración de los mapas perceptuales.

Cuarto paso: *Deducción de los resultados del análisis multidimensional y evaluación de ajuste conjunto*

Se utilizó INDSCAL [4] para representar en un mapa perceptual de composición, o agregado, las medidas de las diferencias de percepción entre los encuestados. Los 45 juicios de similitud de los 18 encuestados fueron introducidos como matrices separadas, pero se calculó una matriz de puntuaciones medias para ilustrar la pauta general de similitudes (véase Tabla 10.4). La tabla también detalla las similitudes más altas (mayores de 6,0) así como la similitud más baja para cada empresa. Con estas relaciones, se puede identificar y disponer de las pautas básicas para compararlas con el mapa resultante.

El primer análisis de los resultados del análisis multidimensional determina la dimensionalidad apropiada y representa los resultados en un mapa perceptual. Para hacerlo, el investigador debería considerar tanto los índices de ajuste en cada dimensionalidad como la capacidad de interpretación de la solución. La Tabla 10.5 muestra los índices de ajuste para soluciones de dos o cinco dimensiones (una solución unidimensional no se consideró una alternativa viable para 10 empresas). Como muestra la tabla, existe una mejora sustancial al pasar de dos a tres dimensiones, después de lo cual las mejoras disminuyen en cierta medida y es consistente a medida que aumentamos en número de dimensiones. Sopesando esta mejora en el ajuste frente a la mayor difi-

TABLA 10.4. Calificaciones de similitud media para HATCO y nueve empresas competidoras

	Empresas									
Empresas	*HATCO*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
HATCO	0,00									
A	6,61	0,00								
B	5,94	5,39	0,00							
C	2,33	2,61	3,44	0,00						
D	2,56	2,56	4,11	6,94	0,00					
E	4,06	2,39	2,17	4,06	2,39	0,00				
F	2,50	3,50	4,00	2,22	2,17	4,06	0,00			
G	2,33	2,39	3,72	2,67	2,61	3,67	2,28	0,00		
H	2,44	4,94	6,61	2,50	7,06	5,61	2,83	2,56	0,00	
I	6,17	6,94	2,83	2,50	2,50	3,50	6,94	2,44	2,39	0,00
Calificaciones máximas y mínimas de similitud para cada empresa										
Similitudes > 6,0	A, I	HATCO I	H	D	C, H	None	I	None	B, D	HATCO A, F
Similitudes mínimas	C, G	E, G	E	F	F	B	C	F	I	H

Nota: las calidicaciones de similitud se realizan sobre una escala de nueve puntos (1 = absolutamente diferentes, 9 = muy parecidos).

TABLA 10.5. Evaluación del ajuste del modelo conjunto y determinación de la dimensionalidad apropiada

Dimensionalidad de la solución	*Medidas medias de ajuste*[a]			
	Stress[b]	*Cambio porcentual*	*R cuadrado*[c]	*Cambio porcentual*
5	0,20068	—	0,6303	—
4	0,21363	6,4	0,5557	11,8
3	0,23655	10,7	0,5007	9,9
2	0,30043	27,0	0,3932	21,5

[a]Media para 18 soluciones individuales.
[b]Fórmula del stress de Kruskal.
[c]Proporción de las calificaciones de similitud original a partir de datos medidos (distancias) desde el mapa percentual.

cultad de interpretación, las soluciones de dos-tres dimensiones parecen las más apropiadas. A efectos ilustrativos, se selecciona la solución de dos dimensiones para análisis ulteriores, pero los métodos que discutimos aquí se pueden aplicar fácilmente a la solución tridimensional. Se anima al investigador a explorar otras soluciones para evaluar si se alcanzarán conclusiones sustancialmente diferentes basándose en la dimensionalidad seleccionada.

El mapa perceptual agregado de dos dimensiones se muestra en la Figura 10.11. HATCO se asocia más con la empresa A, con encuestados que las consideran casi idénticas. Otros pares de empresas consideradas altamente similares basándose en su proximidad son E, G, D y H, y F e I. También se pueden hacer comparaciones entre estas empresas y HATCO. HATCO difiere de C, E y G principalmente en la dimensión II, mientras que la dimensión I opone a HATCO más claramente de las empresas B, C, D y H en una dirección y de las empresas F e I en otra dirección. Todas estas diferencias se reflejan en sus posiciones relativas en el mapa perceptual. Se pueden realizar comparaciones parecidas entre todos los conjuntos de empresas. Para entender las fuentes de estas diferencias, sin embargo, el investigador debe interpretar las dimensiones.

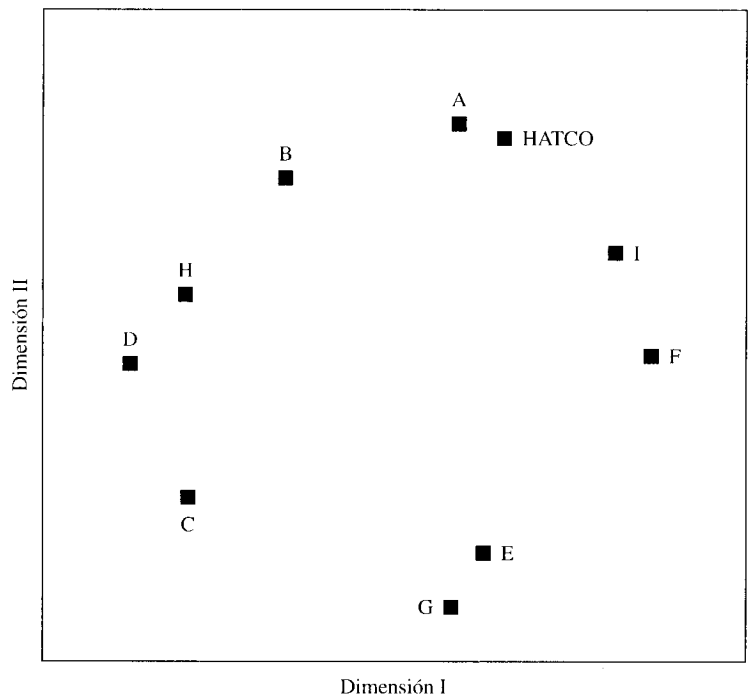


FIGURA 10.11. Mapa perceptual de HATCO y de los principales competidores.

El investigador también puede observar el ajuste de la solución en un gráfico de dispersión de las distancias efectivas (valores de similitud medidos) frente a las distancias ajustadas a partir del mapa perceptual (véase Figura 10.12). Este gráfico puede identificar auténticos casos atípicos que no están bien representados por la solución corriente. Si un conjunto consistente de objetos o individuos se identifica como caso atípico, puede considerarse su eliminación. En este caso, ninguna empresa muestra un número grande de puntos que sea candidato a la eliminación del análisis.

Además de desarrollar el mapa de percepción de composición, INDSCAL también proporciona los medios para evaluar uno de los supuestos del análisis multidimensional, la homogeneidad de las percepciones de los encuestados. Se calculan las ponderaciones para cada encuestado indicando la correspondencia de su propio espacio perceptual y el mapa perceptual agregado. Éstas ponderaciones ofrecen una medida de las comparaciones entre los encuestados, dado que los encuestados con ponderaciones similares tienen el mismo mapa perceptual. INDSCAL también proporciona una medida del ajuste para cada sujeto mediante la correlación de las puntuaciones calculadas y las calificaciones de similitud originales del encuestado. La Tabla 10.6 contiene las ponderaciones y las medidas de ajuste de cada encuestado y la Figura 10.13 es una representación gráfica de los encuestados individuales basándose en sus ponderaciones. El análisis de las ponderaciones (Tabla 10.6) y la Figura 10.12 revelan que los encuestados son bastante homogéneos en sus percepciones, dado que las ponderaciones muestran pocas diferencias sustantivas sobre cada dimensión, y no aparecen «conglomerados» indiferenciados de individuos. Esto se demuestra en la Figura 10.12, ya que todas las ponderaciones de los individuos caen claramente en una línea recta, indicando una ponderación consistente entre las dimensiones I y II. La distancia de cada ponderación individual hasta el origen indica su nivel de ajuste con el de la solución. Se obtienen mejores ajustes para mayores distancias hasta el origen. Por tanto, los encuestados 4, 7 y 10 tienen el mejor ajuste mientras que los encuestados 1 y 9 tienen el peor ajuste. Los valores del ajuste muestran una consistencia relativa tanto en la medida del stress como en la del R^2, con valores medios de 0,003 (stress) y 0,393 (R^2). Es más, todos los encuestados están bien representados por el mapa perceptual de composición, siendo la menor medida del ajuste un 0,27. Por tanto, no debería eliminarse ningún individuo debido al mal ajuste de una solución de dos dimensiones.

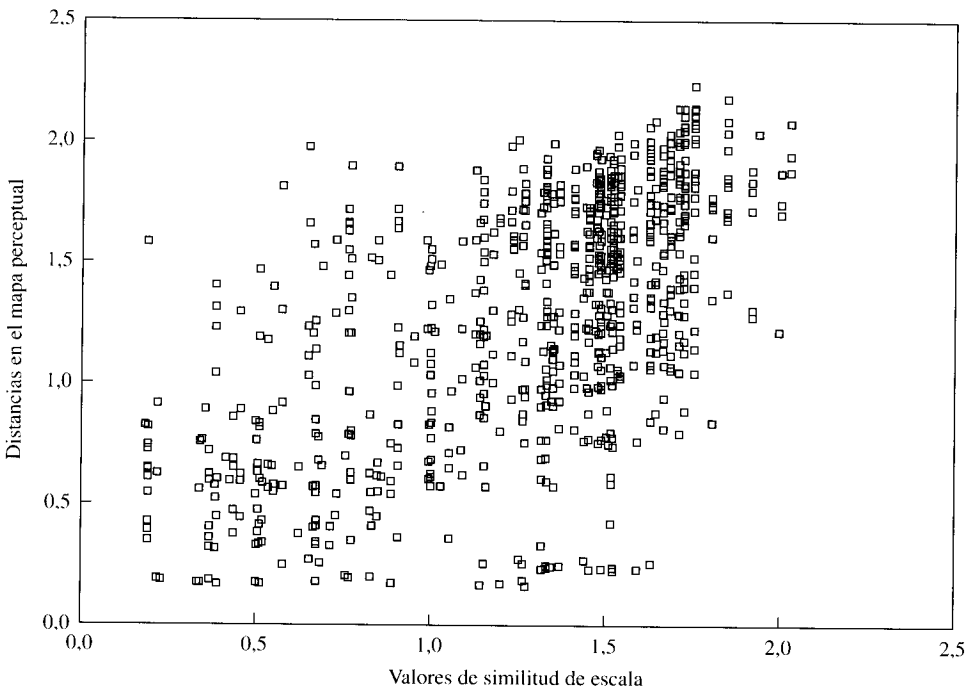


FIGURA 10.12. Gráfico de dispersión de un ajuste lineal.

TABLA 10.6. Medida de las preferencias de los individuos en la elaboración del mapa perceptual: medidas de ajuste específicas de los encuestados y ponderación de las dimensiones

	Medidas de ajuste		*Ponderación de las dimensiones*	
Sujeto	*Stress*[a]	*R cuadrado*[b]	*Dimensión I*	*Dimensión II*
1	0,358	0,274	0,386	0,353
2	0,297	0,353	0,432	0,408
3	0,302	0,378	0,395	0,472
4	0,237	0,588	0,572	0,510
5	0,308	0,308	0,409	0,375
6	0,282	0,450	0,488	0,461
7	0,247	0,547	0,546	0,499
8	0,302	0,332	0,444	0,367
9	0,320	0,271	0,354	0,382
10	0,280	0,535	0,523	0,511
11	0,299	0,341	0,397	0,429
12	0,301	0,343	0,448	0,378
13	0,292	0,455	0,497	0,456
14	0,302	0,328	0,427	0,381
15	0,290	0,371	0,435	0,426
16	0,311	0,327	0,418	0,390
17	0,281	0,433	0,472	0,458
18	0,370	0,443	0,525	0,409
Promedio[c]	0,300	0,393		

[a]Fórmula de *stress* de Kruskal.
[b]Proporción de las calificaciones originales de similitud tenidas en cuenta en los datos medidos a partir del mapa percentual.
[c]Media para las soluciones de 18 individuos.

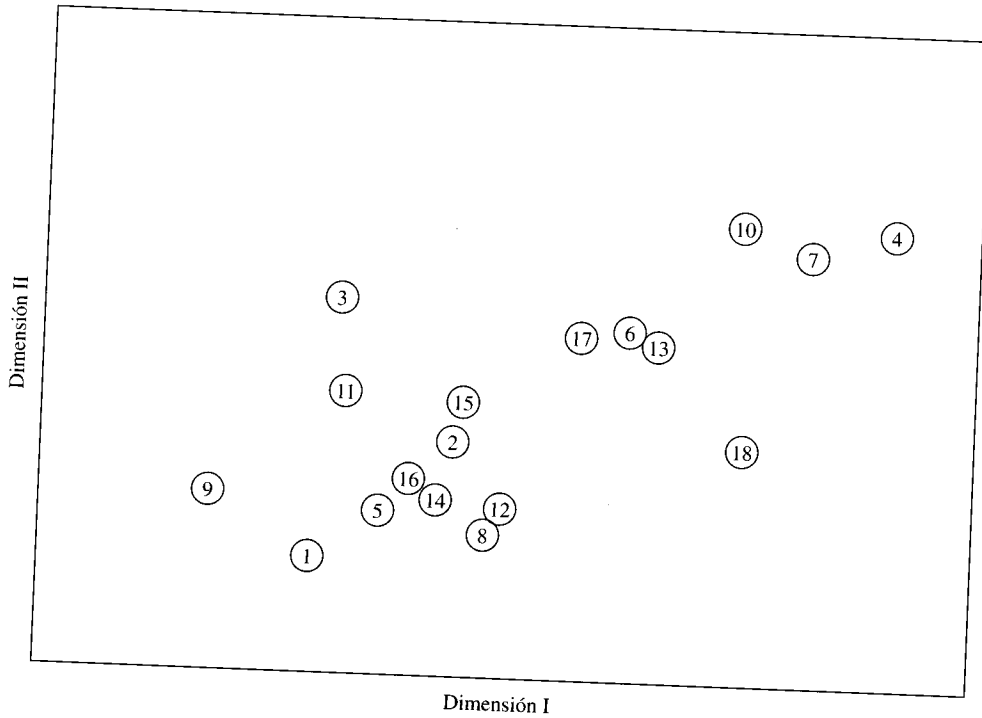


FIGURA 10.13. Ponderaciones de los sujetos individuales.

Incorporación de las preferencias al mapa perceptual Hasta este punto, hemos tratado sólo con los juicios sobre las empresas basados en su similitud, pero muchas veces desearíamos extender el análisis a los procesos de toma de decisiones y entender las preferencias de los encuestados por los objetos (en este caso, empresas). Para hacerlo, podemos emplear varias técnicas de análisis multidimensional que permiten la estimación de los puntos ideales, de los que se pueden determinar las preferencias de los objetos. En este ejemplo, empleamos un método externo de formación de preferencias (prefmap [6]) que utiliza mapas perceptuales agregados derivados en la sección anterior.

Las preferencias fueron medidas pidiendo a los encuestados que detallasen sus preferencias por empresas en diferentes situaciones de compra. Concretamente, aquí examinamos las preferencias por empresas en la situación de compra nueva. Los datos del programa incluyen las coordenadas de las empresas en el mapa agregado perceptual y las preferencias de cinco encuestados. Las calificaciones de preferencia para estos cinco encuestados se dan en la Tabla 10.7.

El programa puede estimar puntos ideales basados en vectores o en puntos. En esta situación, la ejecutiva de HATCO se decantó por las representaciones de puntos, que resultaron en la derivación de puntos ideales para los cinco encuestados, más un punto ideal para el sujeto «medio». Los resultados se muestran en la Figura 10.14. Las distancias de cada empresa a los puntos ideales se ofrecen en la Tabla 10.7. Los valores más bajos indican una mayor proximidad a un punto ideal.

Todos los encuestados forman un grupo general de alguna forma agrupado alrededor de la media. Sin embargo, podemos detectar todavía diferencias en la proximidad al grupo en conjunto así como para las empresas individuales. En primer lugar, el grupo como conjunto, está más cercano a las empresas C, D, F y H, mientras HATCO, A, B, E y G se encuentran de alguna forma más lejos. Nótese que en este caso tanto la dimensionalidad como la proximidad son importantes. El supuesto de un análisis externo es que en la medida en que se cambie su posición en el mapa perceptual sobre las dimensiones, puede cambiar su proximidad a los puntos ideales y su ordenación de preferencias. En términos de los encuestados individuales, hay algunas relaciones cercanas. El encuestado 1 tiene una asociación relativamente estrecha con la empresa F y los encuestados 3 y 5 tienen asociaciones estrechas con las empresas C, D y H. Aunque este grupo de encuestados es relativamente homogéneo en sus preferencias, como indica su agrupamiento conjunto, la Figura 10.13 todavía retrata la posición relativa de cada empresa no sólo en percepción, sino también en preferencia.

TABLA 10.7. Datos de preferencia de las situaciones de nueva compra para encuestados seleccionados

	Empresa										
Sujeto	*HATCO*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*Ajuste*[a]
1	2	3	5	6	7	4	10	8	1	9	
	−0,867	−0,972	−0,920	−1,096	−1,095	−0,636	−0,264	−1,054	−0,854	−0,371	0,787
2	5	2	7	6	9	3	4	1	10	8	
	−1,049	−1,056	−0,622	−0,906	−0,642	−1,111	−0,879	−1,596	−0,413	−0,825	0,961
3	4	1	8	7	6	9	3	5	10	2	
	−0,894	−0,868	−0,448	−0,133	−0,106	−0,449	−0,726	−0,576	−0,132	−0,779	0,855
4	4	3	10	2	7	8	6	1	9	5	
	−1,098	−1,128	−0,736	−1,060	−0,813	−1,136	−0,822	−1,672	−0,544	−0,790	0,884
5	4	1	8	7	9	3	5	2	10	6	
	−0,905	−0,868	−0,401	−0,362	−0,188	−0,769	−0,870	−1,019	−0,126	−0,838	0,977
Promedio	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	
	−0,916	−0,931	−0,580	−0,668	−0,525	−0,776	−0,666	−1,140	−0,370	−0,674	0,990

Nota: los valores de la parte de arriba de cada celda son las calificaciones originales de preferencias, y las entradas de abajo son las distancias al cuadrado de las empresas respecto del punto ideal. NA indica que las calificaciones medias no están disponibles.

[a]El ajuste es la correlación cuadrada entre las preferencias y los valores de las distancias.

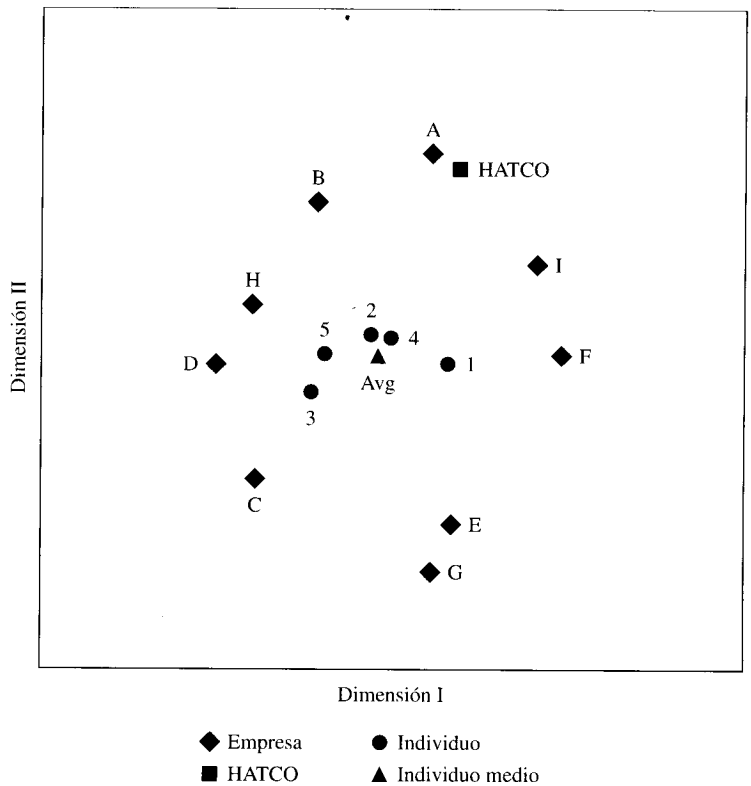
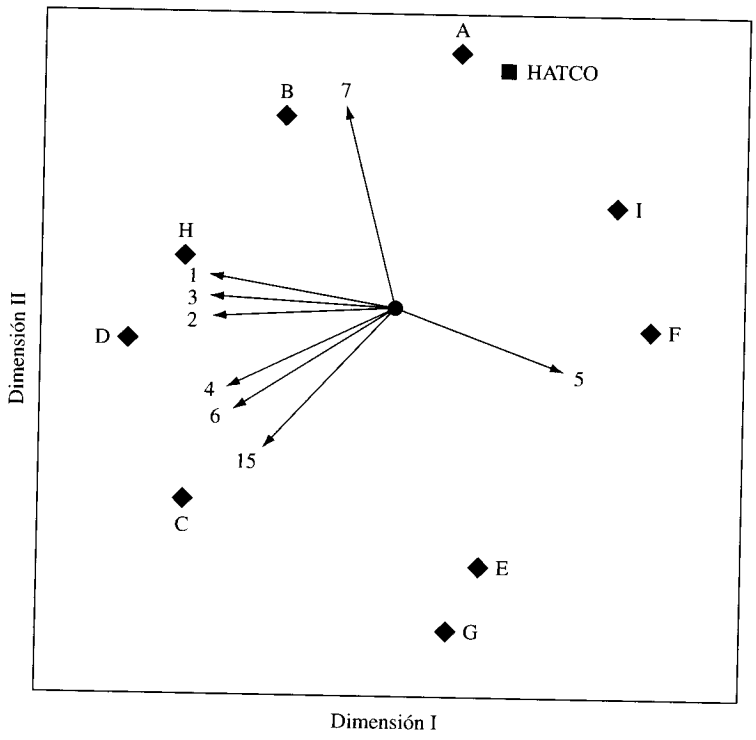


Figura 10.14. Mapa de los puntos ideales para encuestados medios y seleccionados: situación de nueva compra.

Quinto paso: *Interpretación de los resultados*

Una vez que se ha establecido el mapa perceptual, podemos empezar el proceso de interpretación. Dado que el procedimiento INDSCAL utiliza sólo los juicios de similitud conjunta, HATCO también reúne las calificaciones de las empresas en ocho atributos —las siete evaluaciones utilizadas antes y una nueva variable, X_{15}, representando la orientación estratégica— descriptivos de las estrategias típicas seguidas en esta industria. Las calificaciones de cada empresa se promediaron a continuación para los encuestados para una calificación global única. Para ofrecer un medio objetivo de interpretación, se utilizó *profit* [3], un modelo de vector, para captar las calificaciones de las posiciones de las empresas en el mapa perceptual. Los resultados de aplicar los datos de clasificación al mapa perceptual de composición se muestran en la Figura 10.15. Esta figura muestra que existen tres tipos distintos de «grupos» o dimensiones de atributos. El primero comprende X_1 (velocidad de entrega), X_2 (nivel del precio) y X_3 (flexibilidad de precios), que apuntan todas a la misma dirección, y X_5 (servicio conjunto), que está en la dirección opuesta a la de las tres variables orientadas al precio. Esta diferencia de direcciones indica una correspondencia negativa del servicio frente a las tres variables restantes. El segundo conjunto de variables refleja evaluaciones más globales, que consisten en las dos variables de imagen, X_6 (imagen de la fuerza de ventas) y X_4 (imagen del producto), junto con una nueva variable, X_{15} (orientación estratégica). Finalmente, X_7 (calidad del producto) es casi perpendicular con respecto a las dimensiones de precio y servicio, indicando una dimensión de evaluación distinta y separada.

Para interpretar las dimensiones, el investigador busca atributos muy alineados con los ejes. En este caso, los grupos de atributos están ligeramente en ángulo desde los ejes originales. Sin embargo, dado que el mapa perceptual es una representación de punto, los ejes pueden moverse sin que exista un impacto en las posiciones relativas. Si giramos los ejes ligeramente (tanto como se



Empresas:
- HATCO
- Competidores

Atributos:
1. X_1 Velocidad de entrega
2. X_2 Nivel de precio
3. X_3 Flexibilidad de precios
4. X_4 Imagen del producto
5. X_5 Servicio conjunto
6. X_6 Imagen de la fuerza de ventas
7. X_7 Calidad del producto
15. X_{15} Orientación estratégica

FIGURA 10.15. Mapa perceptual con atributos del vector de representación.

hizo en el análisis factorial del Capítulo 3), tendríamos ahora una dimensión de precio y servicio frente a una segunda dimensión de calidad de producto. Aunque no es necesario realizar de forma efectiva la rotación porque las empresas pueden ser comparadas directamente con los vectores de atributos, muchas veces la rotación puede ayudar a la obtención de una comprensión más fundamental de las dimensiones percibidas. La rotación es especialmente útil en las soluciones que comprenden más de dos dimensiones.

Para determinar los valores de cualquier empresa sobre un vector de atributo, necesitamos calcular las proyecciones desde la empresa al vector. Para ayudar en la evaluación, el programa PROFIT ofrece los valores de la proyección para cada atributo, que se enumeran en la segunda columna de valores para cada variable en la Tabla 10.8. También están incluidas las calificaciones originales (valores de la primera columna) para ver lo bien que el vector representa las percepciones efectivas de los encuestados. Por ejemplo, el orden de las calificaciones de las empresas acerca de la calidad del producto (de mayor a menor) era H, B, A, HATCO, F, E, D, C, I y G. Utilizando las proyecciones de los vectores, vemos que el orden de las empresas es, A, B, HATCO, H, D, I, F, C, E y G. Existe una clara correspondencia entre los valores originales y los calculados, particularmente entre las cuatro primeras empresas. Una medida estadística de ajuste para cada atributo es la correlación entre las calificaciones originales y las proyecciones del vector. En el caso de la calidad del producto, la correlación es de 0,710. El investigador no debería esperar un

TABLA 10.8. Interpretación del mapa perceptual con PROFIT

	Calificaciones de atributos originales y proyecciones de los vectores ajustados										
	Empresa										
Variables	*HATCO*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*Ajuste*[a]
X_7 Calidad del producto	6,94	7,17	7,67	3,22	4,78	5,11	6,56	1,61	8,78	3,17	0,710
	1,0038	1,1529	1,1057	−0,6284	0,2474	−1,4128	−0,5099	−1,6815	0,5717	0,1511	
X_{15} Orientación estratégica	4,0	1,83	6,33	7,67	6,00	5,78	5,50	6,11	7,50	4,17	0,785
	−1,2912	−1,1645	−0,2293	1,5350	1,1914	0,5743	−0,9644	0,9370	0,6718	−1,2600	
X_5 Servicio o conjunto	6,94	5,67	3,39	3,67	3,67	6,94	6,44	7,22	4,94	6,11	0,842
	0,1994	−0,0704	−0,9018	−0,8079	−1,3700	0,8879	1,4219	0,8185	−1,2092	1,0317	
X_1 Velocidad de entrega	4,00	3,39	7,33	6,11	7,50	4,22	7,17	4,33	8,22	5,56	0,510
	−0,4202	−0,1535	0,7696	1,0137	1,4560	−0,6928	−1,4604	−0,5627	1,2144	−1,1641	
X_2 Nivel de entrega	5,16	3,47	6,41	5,88	6,06	4,94	5,29	4,82	8,35	4,65	0,653
	−0,5571	−0,2965	0,6725	1,1322	1,4903	−0,5561	−1,4639	−0,3885	1,1998	−1,2328	
X_6 Imagen de la fuerza de ventas	5,11	1,22	5,78	7,89	6,56	3,83	4,28	6,94	8,67	4,72	0,720
	−1,1133	−0,9244	0,1007	1,5053	1,4050	0,1842	−1,2369	0,5012	0,9313	−1,3530	
X_3 Flexibilidad de precios	5,33	3,72	6,33	5,56	6,39	4,72	5,28	5,22	7,33	5,11	0,651
	−0,5138	−0,2509	0,7045	1,0956	1,4812	−0,6007	−1,4666	−0,4450	1,2060	−1,2122	
X_4 Imagen del productor	4,17	1,56	6,06	8,22	7,72	4,28	3,89	6,33	7,72	5,06	0,829
	−1,0905	−0,8962	0,1333	1,4956	1,4201	0,1443	−1,2587	0,4553	0,9531	−1,3563	

Nota: los valores superiores de las celdas son las calificaciones de los atributos originales; los números inferiores on las proyecciones para los valores ajustados.

[a]El ajuste se mide como la correlación entre las calificaciones de los atributos originales y el vector de proyeccioens.

ajuste perfecto por varias razones. En primer lugar, el mapa perceptual está basado en la evaluación conjunta, que puede no ser directamente comparable con la evaluación aislada. En segundo lugar, las calificaciones son promediadas para todos los encuestados, de tal forma que sus valores se determinan por las diferencias entre los individuos, así como las diferencias entre las empresas. Dados estos factores, el nivel de ajuste para los atributos es individual y colectivamente aceptable.

Revisión de los resultados de descomposición

Los métodos de descomposición empleados en este estudio de imagen ilustran tanto las ventajas como las desventajas de esta aproximación. El uso de juicios de similitud, proporciona un mapa perceptual basado sólo en criterios relevantes elegidos por cada encuestado. Sin embargo, las técnicas libres de atributos también demuestran la notable dificultad de interpretar un mapa perceptual en términos de atributos específicos. Exige del investigador la inferencia de bases para la comparación entre objetos sin una directa confirmación para el encuestado.

Los resultados ofrecen una perspectiva de las percepciones relativas de HATCO y las otras nueve empresas. En términos de percepciones, HATCO está más asociada con la empresa A y en cierta forma con las empresas B e I. Existen algunos grupos competitivos (por ejemplo, F e I, E y G), que deben ser considerados. Ninguna de estas empresas es lo suficientemente diferente como para ser considerada un atípico. HATCO puede considerarse «media» en varios atributos (X_1, X_2, y X_3), pero tiene bajas puntuaciones en varios atributos (X_4, X_6, y X_{15}) contando con una alta puntuación sobre el atributo X_7. Finalmente, HATCO no tiene ventajas reales en términos de proximidad a los puntos ideales del encuestado, con otras empresas, tales como D, H y F, estando localizada como más próxima a los puntos ideales de varios encuestados. Estos resultados ofrecen una perspectiva de HATCO no sólo de sus percepciones, sino también de las percepciones de otros principales competidores del mercado.

Análisis de correspondencias: Pasos cuarto y quinto

Una alternativa a la elaboración de mapas perceptuales libres de atributos es CA, un método de composición que se basa en ocho calificaciones binarias de las empresas ofrecidas por cada empresa (es decir, las calificaciones sí—no de cada empresa para cada atributo). En este método basado en los atributos, el mapa perceptual es un espacio conjunto, que muestra tanto atributos como empresas en una única representación. Es más, las posiciones de las empresas son relativas no sólo respecto a otras empresas incluidas en el análisis, sino también respecto a los atributos seleccionados.

Cuarto paso: *Deducción del análisis de correspondencias*

La preparación de los datos para el análisis implica la creación de una matriz de tabulación cruzada que relaciona los atributos (representados como filas) con las calificaciones de las empresas (las columnas). Las entradas individuales de la matriz son el número de veces que se califica a una empresa como poseyendo un atributo específico. Por tanto, se proporcionan frecuencias simples para cada empresa para todo el conjunto de atributos (véase Tabla 10.9).

El análisis de correspondencias se basa en una transformación del valor de la chi-cuadrado en una medida métrica de la distancia. El valor de la chi-cuadrado se calcula como la frecuencia efectiva de ocurrencia menos la frecuencia esperada. Por tanto, un valor negativo indica, en este caso, que una empresa fue asociada con los atributos menos a menudo de lo que se esperaba. El valor esperado de una celda (cualquier combinación atributo-empresa en la tabla de tabulación cruzada) se basa en la frecuencia con que la empresa fue asociada con otros atributos y la frecuencia con que otras empresas se asociaron con el atributo. (En términos estadísticos, el valor esperado se basa en las probabilidades marginales de la fila [atributo] y la columna [empresa].) La Tabla 10.10 contiene las distancias chi-cuadrado (métricas) transformadas para cada celda de tabulación cruzada a partir de la Tabla 10.9. Valores positivos elevados indican un fuerte grado de «correspondencia» entre el atributo y la empresa, y los valores negativos tienen una interpretación opuesta. Por ejemplo, los valores altos de HATCO y las empresas A y F en el atributo imagen del productor (X_4) indican que deberían estar muy próximas en el mapa perceptual, si es que eso es posible. De la misma forma, valores negativos elevados para las empresas C y D sobre la misma variable indicarían que sus posiciones deberían estar alejadas de la localización de atributo.

CA intenta satisfacer todas estas relaciones simultáneamente mediante la producción de dimensiones representativas de las distancias chi-cuadrado. Para determinar la dimensión de la solución, el investigador examina el porcentaje acumulado de variación explicada, igual que en el análisis factorial, y determina la dimensión apropiada. La Tabla 10.11 contiene los autovalores y los porcentajes de variación acumulada y explicada para cada dimensión hasta un máximo de sie-

TABLA 10.9. Datos de frecuencias de tabulación cruzada de atributos que describen a HATCO y a nueve empresas competidoras

	Empresa									
Variables	*HATCO*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
X_7 Calidad del producto	4	3	1	13	9	6	3	18	2	10
X_{15} Orientación estratégica	15	16	15	11	11	14	16	12	14	14
X_5 Servicio conjunto	15	14	6	4	4	15	14	13	7	13
X_1 Velocidad de entrega	16	13	8	13	9	17	15	16	6	12
X_2 Nivel de entrega	14	14	10	11	11	14	12	13	10	14
X_6 Imagen de la fuerza de ventas	7	18	13	4	9	16	14	5	4	16
X_3 Flexibilidad de precios	6	6	14	10	11	8	7	4	14	4
X_4 Imagen delproductor	15	18	9	2	3	15	16	7	8	8

TABLA 10.10. Medidas de similitud en el análisis de correspondencias: distancias Chi-cuadrado

	Empresa									
Variables	*HATCO*	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
X_7 Calidad del producto	−1,27	−1,83	−2,08	3,19	1,53	−0,86	−1,73	4,07	−1,42	0,97
X_{15} Orientación estratégica	0,02	−0,13	0,76	−0,01	0,04	−0,73	0,07	−0,60	1,07	−0,20
X_5 Servicio conjunto	1,08	0,40	−1,10	−1,52	−1,48	0,57	0,59	0,65	−0,36	0,53
X_1 Velocidad de entrega	0,68	−0,51	−0,95	0,95	−0,27	0,40	0,20	0,86	−1,15	−0,37
X_2 Nivel de entrega	0,19	−0,19	−0,30	0,37	0,42	−0,30	−0,54	0,08	0,20	0,23
X_6 Imagen de la fuerza de ventas	−1,32	−1,49	1,15	−1,54	0,23	0,81	0,55	−1,80	−1,44	1,39
X_3 Flexibilidad de precios	−1,02	−1,28	2,37	1,27	1,71	−0,73	−0,83	−1,59	2,99	−1,66
X_4 Imagen del productor	1,24	1,69	−0,01	−2,14	−1,76	0,72	1,32	−1,07	0,10	−0,85

te. De nuevo, el investigador tiene que ponderar el deseado aumento en la explicación con la interpretación. Una solución bidimensional en esta situación explica el 86 por ciento de la variación, mientras que un aumento a una solución de tres dimensiones añade sólo un 10 por ciento adicional. El investigador debe comparar y contrastar la variación adicional explicada en relación con el aumento de la complejidad en la interpretación de los resultados. Por tanto, se juzga que una solución bidimensional es la más adecuada para el análisis ulterior.

Quinto paso: *Interpretación de los resultados del análisis de correspondencias*

El mapa perceptual basado en atributos, muestra las proximidades relativas tanto de las empresas como de los atributos (véase Figura 10.16). Si nos centramos primero en las empresas, vemos que la pauta de los grupos de empresas es similar a la que muestran los resultados del análisis multidimensional. Las empresas A, E, F e I, más HATCO, forman un grupo; las empresas C y D y las empresas H y B forman otros dos grupos similares. Sin embargo, las proximidades relativas de los miembros en cada grupo difieren en cierta medida de la solución multidimensional. De la misma forma, la empresa G está más aislada y separada, y las empresas F y E son ahora más parecidas a HATCO.

Puede ser de utilidad interpretar las dimensiones si se utilizan normalizaciones de filas y columnas. A estos efectos, la inercia (variación explicada) de cada dimensión puede atribuirse a las filas y columnas. La Tabla 10.12 ofrece las contribuciones de ambos conjuntos de categorías para cada dimensión. Para los atributos, podemos ver que X_7 (calidad del producto) es el que más contribuye a la dimensión I (mayores o menores valores en la dimensión I) y X_4 (imagen del fabricante) sólo contribuye secundariamente. Nótese que existen dos atributos extremos en términos de su localización en la dimensión I (es decir, mayores o menores valores en la dimensión I). Entre estos dos atributos, se tiene en cuenta el 86 por ciento de la dimensión I. Se sigue una pauta similar para

TABLA 10.11. Determinación de la dimensión apropiada para el análisis de correspondencias

Dimensión	*Autovalor (Valor singular)*	*Inercia (Chi-cuadrado normalizado)*	*Porcentaje Explicado*	*Porcentaje Acumulado*
1	0,27666	0,07654	53,1	53,1
2	0,21866	0,04781	33,2	86,3
3	0,12366	0,01529	10,6	96,9
4	0,05155	0,00266	1,8	98,8
5	0,02838	0,00081	0,6	99,3
6	0,02400	0,00058	0,4	99,7
7	0,01951	0,00038	0,3	100,0

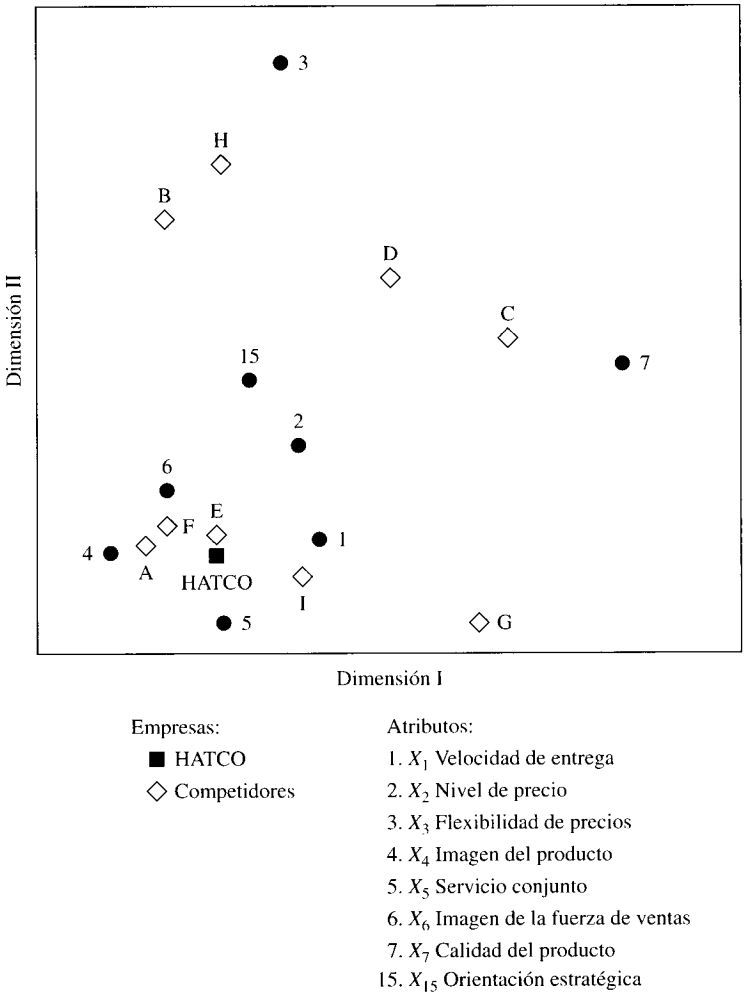


Figura 10.16. Elaboración de mapa perceptual con métodos de composición: análisis de correspondencias.

la dimensión II, para la cual X_3 (flexibilidad de precios) es el principal componente seguido por X_5 (servicio conjunto). Si miramos a las 10 empresas, vemos que de alguna forma la situación está compensada, con tres empresas (A, C y G) contribuyendo en una media por encima del 10 por ciento. Para la segunda dimensión, cuatro empresas (B, D, G y H) tienen contribuciones por encima de la media. Aunque las comparaciones de este ejemplo se encuentran entre ambos conjuntos de categorías y no restringidas a un único conjunto de las mismas (o fila o columna), estas medidas de contribución demuestran la capacidad para interpretar la dimensión cuando se desee.

Hay una última medida que nos ofrece una evaluación del ajuste para cada categoría. Comparable a las cargas de los factores al cuadrado del análisis factorial (véase Capítulo 3 para una discusión más detallada), estos valores representan la cantidad de la variación en la categoría explicada por la dimensión. Un valor total representa la cantidad total de variación para todas las dimensiones, siendo el máximo posible el 100 por cien. La Tabla 10.12 contiene los valores del ajuste para cada categoría en cada dimensión. Como podemos ver, los valores del ajuste van desde un máximo de 99,1 para X_1 (velocidad de entrega) y X_7 (calidad del producto) a un mínimo de 0,372 para X_6 (imagen de la fuerza de ventas). Entre los atributos, sólo X_6 (imagen de la fuerza de ventas) tiene un valor por debajo del 50 por ciento y sólo dos empresas (HATCO y empresa I) caen por debajo de este valor. Incluso aunque éstos sean bajos, todavía representan una explicación suficiente para retenerla en el análisis y juzgar el análisis de suficiente significación práctica.

TABLA 10.12. Interpretación de las dimensiones y sus correspondencias para empresas y atributos

	Coordenadas		*Contribución a la inercia*[a]		*Explicación por la dimensión*[b]		
Objecto	*I*	*II*	*I*	*II*	*I*	*II*	*Total*
Atributos							
X_7 Calidad del producto	0,044	1,235	0,001	0,689	0,002	0,989	0,991
X_{15} Orientación estratégica	−0,676	−0,285	0,196	0,044	0,789	0,111	0,901
X_5 Servicio conjunto	0,115	0,046	0,007	0,001	0,469	0,058	0,527
X_1 Velocidad de entrega	1,506	0,298	0,665	0,033	0,961	0,030	0,991
X_2 Nivel de precio	−0,081	0,245	0,004	0,045	0,093	0,678	0,772
X_6 Imagen de la fuerza de ventas	−0,440	−0,099	0,087	0,006	0,358	0,014	0,372
X_3 Flexibilidad del precio	−0,202	−0,502	0,018	0,142	0,138	0,677	0,816
X_4 Imagen del productor	0,204	−0,245	0,022	0,040	0,289	0,330	0,619
Empresas							
HATCO	−0,247	−0,293	0,024	0,042	0,206	0,228	0,433
A	−0,537	−0,271	0,125	0,040	0,772	0,156	0,928
B	−0,444	0,740	0,063	0,224	0,294	0,648	0,942
C	1,017	0,371	0,299	0,050	0,882	0,093	0,975
D	0,510	0,556	0,074	0,111	0,445	0,418	0,863
E	−0,237	−0,235	0,025	0,031	0,456	0,356	0,812
F	−0,441	−0,209	0,080	0,023	0,810	0,144	0,954
G	0,884	−0,511	0,292	0,123	0,762	0,201	0,963
H	−0,206	0,909	0,012	0,289	0,049	0,748	0,797
I	0,123	−0,367	0,006	0,066	0,055	0,390	0,446

[a]Proporción de la inercia de la dimensión atribuible a cada categoría.
[b]Proporción de la variación de la categoría tenida explicada por la dimensión.

Resumen de CA

Estas y otras comparaciones resaltan las diferencias entre el análisis multidimensional y el análisis de correspondencias. Esta última técnica proporciona un medio de comparación directa de similitud y diferencia de empresas y de los atributos asociados, mientras que el análisis multidimensional sólo permite las comparaciones entre empresas. Pero la solución CA está condicionada por el conjunto de atributos incluidos. Supone que todos los atributos son apropiados para todas las empresas y que se aplica la misma dimensión para cada empresa. Por tanto, el mapa perceptual resultante debería ser visto siempre sólo en el contexto de las empresas y los atributos incluidos en el análisis.

El análisis de correspondencias es una técnica muy flexible y aplicable a un amplio abanico de supuestos y situaciones. Las ventajas del gráfico conjunto de atributos y objetos debe ponderarse siempre contra las interdependencias inherentes que existen y los potenciales sesgos de un único atributo —o empresa— inapropiado, o lo que es más importante, el atributo omitido de una empresa. Sin embargo, el análisis de correspondencias sigue proporcionando una poderosa herramienta para conseguir una mejor perspectiva para la dirección de la posición relativa de las empresas y de los atributos asociados con esas posiciones.

Sexto paso: Validación de los resultados

Quizá la validación interna más fuerte de este análisis sea evaluar la convergencia entre los resultados de las técnicas de composición y descomposición. Cada técnica emplea diferentes tipos de respuestas del consumidor, pero los mapas perceptuales resultantes son representaciones del mismo espacio perceptual y deberían corresponderse. Si la correspondencia es elevada, el investigador puede pensar que los resultados reflejan el problema tal y como se ha diseñado. El investigador

debería notar que este tipo de convergencia no garantiza la generalización de los resultados a otros objetos o muestras de la población.

La comparación de los métodos de composición y descomposición, mostrados en las Figuras 10.11 y 10.16, puede llevarse a cabo de dos formas distintas: examinando las posiciones relativas de los objetos e interpretando los ejes. Comencemos examinando la posición de las empresas. Cuando las Figuras 10.11 y 10.16 se giran para obtener la misma perspectiva, muestran pautas muy similares, reflejando dos grupos de empresas: empresas B, H, D y C frente a las empresas E, F, G e I. Mientras que las distancias relativas entre las empresas varían entre los dos mapas perceptuales, todavía vemos a HATCO fuertemente asociado con las empresas A e I en cada mapa perceptual. El análisis de correspondencias produce mayor distinción entre las empresas, pero su objetivo es definir las posiciones de las empresas como resultado de las diferencias; por tanto, generará más distinción en sus mapas perceptuales.

La interpretación de los ejes y las características distintivas también muestra pautas parecidas en los dos mapas perceptuales. Para el método de descomposición mostrado en la Figura 10.15, ya anotamos en la anterior discusión que girando los ejes obtendríamos una interpretación más clara. Si giramos los ejes, entonces la dimensión I se asocia con la velocidad de entrega, nivel de precio y flexibilidad del precio (X_1, X_2 y X_3), mientras que la dimensión II refleja la calidad del producto (X_7). Los atributos restantes no están fuertemente asociados con cada eje. En el análisis de correspondencia (Figura 10.15), vemos que existen en realidad tres grupos de atributos representando las dimensiones de evaluación: calidad de producto (X_7), flexibilidad del precio (X_3), y el resto de los atributos en el tercer grupo. Esto es bastante favorable comparado con los resultados del análisis de descomposición, excepto en el caso de la flexibilidad de precios, que no se diferencia como una dimensión separada.

En conjunto, aunque existen algunas diferencias, debido a las características de cada aproximación, la convergencia de los dos resultados arroja cierta validez interna de los mapas perceptuales. Las diferencias de percepción pueden existir para unos pocos atributos, pero las pautas conjuntas de las posiciones de las empresas y las dimensiones de evaluación se ven apoyadas por ambas aproximaciones. La disparidad en el atributo de la flexibilidad de precios ilustra las diferencias entre ambas aproximaciones. El método de descomposición determina la posición basándose en juicios globales, con los atributos aplicados sólo como un intento de explicar las posiciones. El método de composición ubica a la empresa de acuerdo con el conjunto seleccionado de atributos, creando por tanto posiciones basadas en esos atributos. Es más, cada atributo se pondera de la misma forma, distorsionando potencialmente el mapa con atributos irrelevantes. Estas diferencias no lo hacen óptimo o mejor, sino que deben ser entendidas por el investigador para asegurar la selección del método más apropiado para los objetivos del investigador.

Un resumen para la dirección de los resultados del análisis multidimensional

La elaboración de un mapa perceptual es la única técnica que ofrece comparaciones conjuntas que no están al alcance de otro método multivariante. Como tal, sus resultados presentan un amplio rango de perspectivas para el uso de la dirección. La aplicación más común de los mapas perceptuales es la evaluación de la imagen de la empresa para cualquier empresa o grupo de empresas. Como variable estratégica, la imagen puede ser un indicador esencial de la presencia o posición de mercado. En este estudio, encontramos que HATCO está más asociada con las empresas A e I y se diferencia más de las empresas C, E y G. Por tanto, aunque sirve a los mismos mercados, HATCO puede identificar aquellas empresas consideradas similares (diferentes) a su imagen. Los resultados basados no en cualquier conjunto específico de atributos, sino en los juicios generales de los encuestados, presentan la ventaja de no estar sujetos a los juicios subjetivos del investigador en los atributos incluidos o en cómo se ponderan los atributos individuales. Esto está en el «espíritu» de la evaluación de la imagen. Sin embargo, las tecnologías del aná-

lisis multidimensional son menos útiles como guía estratégica, dado que son menos útiles a la hora de prescribir cómo se produce un cambio en la imagen. Las respuestas globales que han sido ventajosas para la comparación trabajan ahora en contra de su explicación. Aunque las técnicas del análisis multidimensional pueden aumentar la explicación de los mapas perceptuales, deben considerarse suplementarias y por tanto esperarse mayores inconsistencias que si se incorporan al proceso. Por tanto, la investigación adicional puede ayudar a «explicar» las posiciones relativas.

Con este fin, los resultados del análisis de correspondencias representan una aproximación de compromiso al intentar configurar los mapas perceptuales desde una perspectiva de composición. La comparación de los resultados del análisis de correspondencias con los de la solución clásica del análisis multidimensional revela cierto número de consistencias, pero también discrepancias. Estas últimas son de esperar, dado que son técnicas basadas en aproximaciones distintas. Lo que queremos ver en estas dos soluciones es alguna pauta general de asociaciones entre empresas (tal como HATCO y las empresas A e I) y entre atributos, particularmente el inconfundible X_3 (flexibilidad de precios) y X_7 (calidad del producto). La directiva de HATCO puede utilizar estos resultados no sólo como guía para una política general, sino también como esquema para una investigación ulterior con otras técnicas multivariantes en cuestiones más específicas.

Resumen

El análisis multidimensional es un conjunto de procedimientos que pueden utilizarse para desplegar las relaciones que representan similitud o preferencia y que se encuentran ocultas en los datos. Se han utilizado con éxito: (1) para identificar segmentos de mercado basándose en juicios de preferencia, (2) para determinar qué productos son más competitivos entre sí (es decir, son más parecidos), y (3) deducir qué criterios utiliza la gente cuando juzga los objetos (es decir, productos, compañías, campañas de publicidad). El análisis multidimensional revela relaciones que parecen oscurecerse cuando se examinan sólo las cifras que resultan del estudio. Un mapa de percepción visual resalta más las relaciones entre los estímulos objeto de estudio. Sin embargo, ha que tener un gran cuidado cuando se intenta utilizar esta técnica, siendo su mal uso algo habitual. El investigador debe familiarizarse mucho con la técnica antes de utilizarla y debería ver el resultado sólo como un primer paso en la determinación de la información perceptual.

Cuestiones de revisión

1. ¿En qué difiere el análisis multidimensional de otras técnicas de interdependencia (análisis cluster o factorial)?
2. ¿Cuál es la diferencia entre los datos de preferencia y similitud, y qué impacto tienen sobre los resultados las técnicas de análisis multidimensional?
3. ¿Cómo se utilizan los puntos ideales en los procedimientos del análisis multidimensional?
4. ¿Cómo difieren los procedimientos métricos y no métricos del análisis multidimensional?
5. ¿Cómo puede el investigador determinar cuándo el análisis multidimensional ha obtenido la mejor solución?
6. ¿Cómo identifica el investigador las dimensiones en el análisis multidimensional? Compare este procedimiento con el análisis factorial.
7. Compare y contraste las técnicas del análisis de correspondencias y del análisis multidimensional.
8. Describa cómo se obtiene la «correspondencia» o asociación desde una tabla de contingencia.

Bibliografía

1. Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer (1986), «Interpoint Distance Comparisons in Correspondence Analysis.» *Journal of Marketing Research* 23 (August): 271-80.
2. Carroll, J. Douglas, Paul E. Green, and Catherine M. Schaffer (1987), «Comparing Interpoint Distances in Correspondence Analysis: A Clarification.» *Journal of Marketing Research* 24 (November): 445-50.
3. Chang, J. J., and J. Douglas Carroll (1968), «How to Use PROFIT, a Computer Program for Property Fitting by Optimizing Nonlinear and Linear Correlation.» Unpublished paper, Bell Laboratories, Murray Hill, N.J.
4. Chang, J. J., and J. Douglas Carroll (1969), «How to use INDSCAL, a Computer Program for Canonical Decomposition of *n*-way Tables and Individual Differences in Multidimensional Scaling.» Unpublished paper, Bell Laboratories, Murray Hill, N.J.
5. Chang, J. J., and J. Douglas Carroll (1969), «How to Use MDPREF, a Computer Program for Multidimensional Analysis of Preference Data.» Unpublished paper, Bell Laboratories, Murray Hill, N.J.
6. Chang, J. J., and J. Douglas Carroll (1972), «How to Use PREFMAP and PREFMAP2—Programs Which Relate Preference Data to Multidimensional Scaling Solution.» Unpublished paper, Bell Laboratories, Murray Hill, N.J.
7. Green, P. E. (1975), «On the Robustness of Multidimensional Scaling Techniques.» *Journal of Marketing Research* 12 (February): 73-81.
8. Green, P. E., and F. Carmone (1969), «Multidimensional Scaling: An Introduction and Comparison of Nonmetric Unfolding Techniques.» *Journal of Marketing Research* 7 (August): 33-41.
9. Green, P. E., F. Carmone, and Scott M. Smith (1989), *Multidimensional Scaling: Concept and Applications.* Boston: Allyn & Bacon.
10. Green, P. E., and Vithala Rao (1972), *Applied Multidimensional Scaling.* New York: Holt, Rinehart and Winston.
11. Greenacre, Michael J. (1984), *Theory and Applications of Correspondence Analyses.* London: Academic Press.
12. Greenacre, Michael J. (1989), «The Carroll-Grenn-Schaffer Scaling in Correspondence Analysis: A Theoretical and Empirical Appraisal.» *Journal of Marketing Research* 26 (August): 358-65.
13. Hoffman, Donna L., and George R. Franke (1986), «Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research.» *Journal of Marketing Research* 23 (August): 213-27.
14. Holbrook, Morris B., William L. Moore, and Russell S. Winer (1982), «Constructing Joint Spaces from Pick-Any Data: A New Tool for Consumer Analysis.» *Journal of Consumer Research* 9 (June): 99-105.
15. Levine, Joel H. (1979), «Joint-Space Analysis of "Pick-Any" Data: Analysis of Choices from an Unconstrained Set of Alternatives.» *Psychometrika* 44 (March): 85-92.
16. Lingoes, James C. (1972), *Geometric Representations of Relational Data.* Ann Arbor, Mich.: Mathesis Press.
17. Kruskal, Joseph B., and Frank J. Carmone (1967), «How to Use MDSCAL. Version 5-M, and Other Useful Information.» Unpublished paper, Bell Laboratories, Murray Hill, N.J.
18. Kruskal, Joseph B., and Myron Wish (1978), *Multidimensional Scaling.* Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-011, Beverly Hills, Calif.: Sage.
19. Lebart, Ludovic, Alain Morineau, and Kenneth M. Warwick (1984), *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices.* New York: Wiley.
20. Maholtra, Naresh (1987), «Validity and Structural Reliability of Multidimensional Scaling.» *Journal of Marketing Research* 24 (May): 164-73.
21. Market ACTION Research Software, Inc. (1989), MAPWISE: *Perceptual Mapping Software.* Peoria, Ill.: Business Technology Center, Bradley University.
22. Raymond, Charles (1974), *The Art of Using Science in Marketing.* New York: Harper & Row.
23. Schiffman, Susan S., M. Lance Reynolds, and Forrest W. Young (1981), *Introduction to Multidimensional Scaling.* New York: Academic Press.
24. Smith, Scott M. (1989), PC-*MDS: A Multidimensional Statistics Package.* Provo, Utah: Brigham Young University.
25. Srinivasan, V., and A. D. Schocker (1973), «Linear Programming Techniques for Multidimensional Analysis of Preferences.» *Psychometrica* 38 (September): 337-69.

LECTURA COMPLEMENTARIA

Aplicación del análisis multidimensional y del análisis factorial de correspondencias*

ENRIQUE BIGNÉ ALCAÑIZ
Universitat Jaume I de Castellón

NATALIA VILA LÓPEZ
Universitat de Valencia

Para ilustrar con fines exclusivamente pedagógicos la utilización de diversas técnicas de análisis de datos examinadas en el presente capítulo, se han realizado 100 entrevistas personales a amas de casa de una comunidad autónoma española, pertenecientes a una población de unos 10.000 habitantes pero con un gran dinamismo (crecimiento poblacional del 20,57 por ciento entre 1981-1993). En esta zona, la densidad es de 42,2 (inferior a la media nacional de 78,1 habitantes por km^2); el nivel de renta gira en torno a las 800.000 pta./habitante y el nivel económico ronda el millón. El estudio se ha ampliado encuestando a otras 100 amas de casa de otra población para analizar diferencias/preferencias entre segmentos. Esta segunda localidad reúne características similares, si bien su población se duplica. Los datos obtenidos permiten desarrollar la aplicación del análisis multidimensional, el análisis factorial de correspondencias, el análisis de preferencias y el análisis de propiedades, para medir el posicionamiento a través de mapas perceptuales de cada uno de los cinco puntos de venta que compiten en ambas localidades: mercado, tienda tradicional, un hipermercado, un supermercado y un gran supermercado[1]. En el desarrollo se seguirán las fases expuestas con anterioridad.

Fase 1: Objetivos de los mapas perceptuales

La finalidad del presente ejemplo es analizar el posicionamiento y las preferencias de las amas de casa en torno a cinco puntos de venta cuyo componente principal es la distribución de productos de alimentación en diferentes formatos comerciales.

Los datos se analizarán sucesivamente en tres fases: (1) con el fin de identificar el posicionamiento en un mapa perceptual, se empleará el análisis multidimensional y el análisis factorial de correspondencias, (2)para obtener las preferencias de los entrevistados hacia dichos establecimientos, se aplicará un análisis de preferencias, (3)por último, para añadir las propiedades en las que destacan cada uno de ellos, se realizará un análisis de ajuste de propiedades (PROFIT).

Fase 2: Diseño de la investigación

El estudio de posicionamiento se inicia con el diseño de la metodología común a toda investigación de mercados. En consecuencia se diseñó un estudio que velaba por la consecución de los objetivos del mismo adoptándose criterios de representatividad de la muestra, selección de los individuos objeto de la entrevista, supervisión del trabajo de campo y similares. Las peculiaridades del estudio relativas a la obtención de datos susceptibles de tratamiento para la obtención de mapas perceptuales fueron las siguientes:

1. Inclusión de una pregunta relativa a la obtención de información sobre el grado de similitud/disimilitud entre pares de establecimientos. Para ello, los entrevistados deben valorar en una escala, las diferentes parejas objeto de comparación. En el presente estudio la pregunta se formuló del siguiente modo: a continuación le mencionaré pares de establecimientos. Por favor, señale el grado de diferencia (disimilitud) entre cada par. Si considera que los dos establecimientos son totalmente diferentes (disimilares) ponga un 7 y un 1 si piensa que uno y otro apenas difieren. Utilice cualquier valor entre 1 y 7 para reflejar el grado de diferencia (disimilitud) entre cada pareja.

[1] La presente aplicación se presenta únicamente con fines pedagógicos. En consecuencia, los datos no tienen por qué representar exactamente la realidad, ni los establecimientos considerados, competir en algún mercado.

* Esta aplicación ha sido específicamente para este libro. Por tanto, manifestamos nuestro agradecimiento a sus autores. (N. de la R. T.)

2. Adicionalmente a la pregunta anterior, en la cual no se le solicita a la entrevistada ninguna valoración sobre las características específicas de cada establecimiento, se introdujo otra pregunta en la cual las amas de casa entrevistadas, sí tenían que valorar cada establecimiento sobre 22 características, propuestas por algunas de ellas en una fase previa cualitativa como aspectos clave en la elección del establecimiento. En este caso, la pregunta debe permitir obtener una matriz de frecuencias que identifique la presencia o ausencia de cada característica en cada establecimiento. En el presente ejemplo la pregunta se formuló en los siguientes términos: a continuación le iré leyendo una serie de características que usted puede pensar que tienen los establecimientos que le voy a indicar. Por favor, dígame, en su opinión qué establecimientos la poseen, teniendo presente que la característica señalada la puede poseer uno, varios o ninguno de los establecimientos. (*Entrevistador*: mostrar TARJETA de establecimientos.)
3. Para la obtención de los datos de preferencias acerca de cada uno de los establecimientos, se precisa un ranking o jerarquía de preferencias. Ello motiva la inclusión de una pregunta que ordene los establecimientos según la preferencia del entrevistado, que en este estudio fue la siguiente: basándose en el conocimiento que usted tiene de los establecimientos, aunque no sea muy profundo, le ruego que los ordene según los prefiera usted de más a menos. (*Entrevistador*: mostrar TARJETA de establecimientos.)
4. Para la realización del análisis PROFIT, se precisa conocer la puntuación promedio que cada establecimiento recibe para cada una de las propiedades/atributos que se desea analizar. Para tal fin, cada encuestado debe evaluar cada establecimiento sobre cada propiedad utilizando una escala por ejemplo que oscile entre 1 y 10 puntos. Así por ejemplo, se podría preguntar el encuestado que señale el grado en que piensa que una propiedad está presente en cada establecimiento, poniendo un 1 si opina que la propiedad no define en absoluto al establecimiento y un 10 si opina que lo define completamente. El encuestado puede graduar su respuesta entre ambos valores extremos.

Fase 3: Consideraciones generales

Resulta evidente que en la realización de los análisis de datos para la obtención de mapas perceptuales, bien sean éstos de similitud o bien de preferencia, es ineludible realizar una adecuada selección de los elementos objeto de comparación. Dicha elección debe basarse en los objetivos de la investigación. Así, si el objetivo fuera el análisis del conjunto de bebidas refrescantes, la cantidad y variedad de marcas o productos sería bien diferente en comparación con otro hipotético estudio centrado únicamente en las bebidas con cola o bajas en calorías. En cualquier caso el criterio aconsejable es incluir aquellas que realmente compiten desde la perspectiva del consumidor.

Una precisión adicional hace referencia a que la información debe obtenerse siempre por comparación de objetos, de modo tal que el análisis de datos posterior refleje el posicionamiento competitivo de los mismos. Como consecuencia de ello debe prestarse especial atención al grado de conocimiento de los distintos elementos objeto de comparación.

I. ANÁLISIS MULTIDIMENSIONAL

Para un único segmento: Fases 4 y 5

Fase 4: Resultados y ajuste del análisis multidimensional

4.a) Input del análisis

Con el presente ejemplo se pretende medir el posicionamiento de cada uno de los cinco puntos de venta que compiten en una localidad: mercado, tienda tradicional, un gran supermercado, un supermercado y un hipermercado.

Para tal fin, contamos con la información facilitada por cien amas de casa de una localidad, en la que operan los 5 puntos de venta mencionados. De la comparación entre pares de establecimientos, se obtiene el input del análisis que es una matriz cuadrada simétrica (Tabla 10.L.1), en cuyas celdas se indica la distancia promedio entre los competidores estudiados tomados por parejas

En esta matriz se observa la poca distancia, o gran similitud o parecido, que existe entre el supermercado y el gran

TABLA 10.L.1. Distancias promedio entre puntos de venta

	Mercado	*T. Tradici.*	*Gran Supermercado*	*Supermercado*	*Hipermrecado*
Mercado	0	3,04	5,22	5,04	5,37
T. Tradici.	3,04	0	4,83	4,75	5,57
Gran Supermercado	5,22	4,83	0	2,59	3,79
Supermercado	5,04	4,75	2,59	0	3,91
Hipermercado	5,37	5,57	3,79	3,91	0

Nota: escala 1-7, donde 1= nada diferentes y 7= muy diferentes

TABLA 10.L.2. Medidas de la bondad del ajuste del modelo

nº dimensiones	*S-Stres*	*RSQ (Coeficiente correlación)*	*STRESS*
1	0,0050	0,99998	0,00225
2	0,0000	0,99999	0,0000

supermercado (2,59) y la gran distancia que guardan el mercado y la tienda tradicional con el resto de puntos de venta, especialmente con el hipermercado.

4.b) Resultados

El programa SPSS genera diversas salidas, entre las cuales se muestran a continuación: los índices de ajuste (S-Stress, RSQ y Stress), las coordenadas de los establecimientos en el mapa perceptual y diversos gráficos de ajuste.

Índices de ajuste:

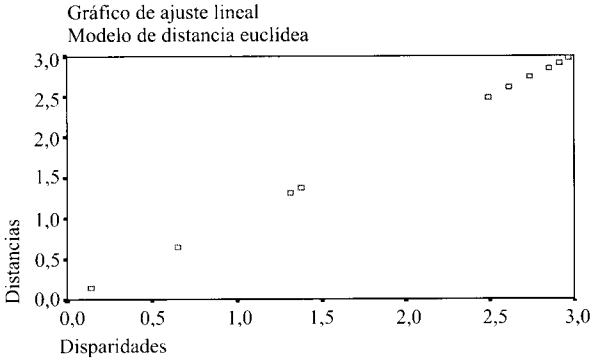
Analizando los distintos indicadores recogidos en la Tabla 10.L.2, se advierte que conforme aumenta el número de dimensiones de 1 a 2, se incrementa la bondad del ajuste[2]. Concretamente, el S-Stress y el Stress son cada vez más próximos a cero y el RSQ se acerca a la unidad. Ello indica que las disimilitudes entre parejas de competidores recogidas en los datos de partida (Tabla 10.L.1) son aproximadas de forma muy exacta por las distancias del mapa perceptual que los representa.

Coordenadas de los establecimientos en el mapa perceptual:

Las coordenadas obtenidas para cada establecimiento sobre cada dimensión muestran su posición respecto a los ejes.

Gráficos:

A continuación, se muestran los gráficos de ajuste (lineal y no lineal) en los que se refleja en qué medida las distancias



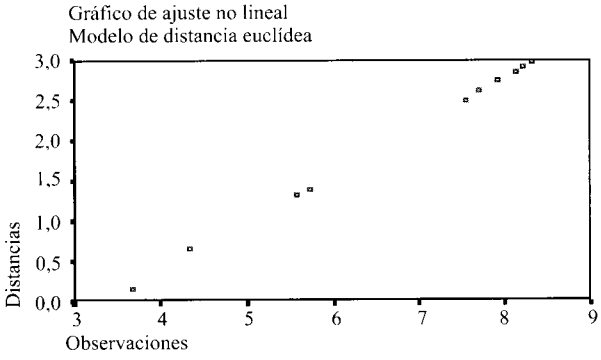


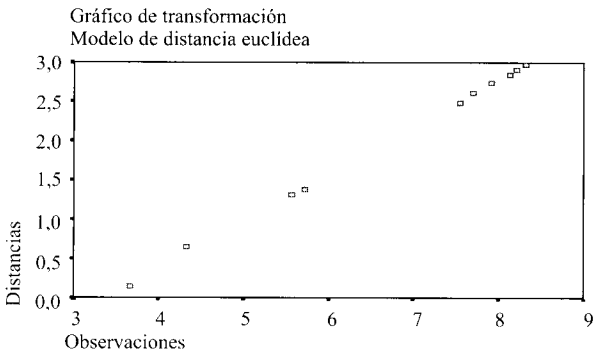
TABLA 10.L.3. Coordenadas en la solución bidimensional

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	1,6052	–0,5341
Tienda tradicional	1,5162	0,3842
Supermercado	–0,9619	0,5222
Gran Supermercado	–0,8067	0,5006
Hipermercado	–1,3528	–0,8730

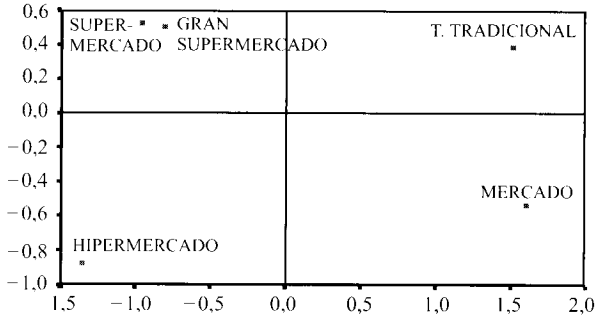
[2] Se podría seguir aumentando el número de dimensiones hasta tres, cuatro, etc. No obstante, dado que en nuestro ejemplo son sólo 5 los competidores objeto de estudio, no tiene mayor sentido continuar haciéndolo.

en el mapa 1, aproximan las disimilitudes entre parejas de establecimientos contenidas en la Tabla 10.L.1.

Como se observa en el gráfico anterior, las distancias euclídeas entre puntos (mapa 1), aproximan con bastante exac-



titud las disimilitudes entre parejas de establecimientos (Tabla 10.L.1). No existen competidores mal representados en la solución final ya que no se observan *outliers* o puntos separados en los gráficos de ajuste. Por ello, no se aconseja eliminar ningún establecimiento del análisis.



MAPA 1. Posicionamiento competitivo de 5 puntos de venta.

Fase 5: Interpretación de resultados

Una vez obtenido el mapa perceptual, se está en condiciones de interpretar los resultados. En primer lugar puede analizarse la semejanza entre pares de puntos de venta en función de la distancia entre ellos. Así, a menor distancia mayor similitud, mientras que a mayor distancia, menor semejanza. El gran supermercado y el supermercado son percibidos de manera muy similar mientras que entre estos dos establecimientos y el resto existe una gran distancia.

En segundo lugar, y complementario al anterior, debe procederse al análisis en función de las coordenadas respecto a los ejes. Analizando el eje horizontal de este mapa se advierte que los puntos de venta que se sitúan en la dimensión negativa, representan fórmulas comerciales más modernas basadas en el autoservicio que las que se sitúan en la dimensión positiva, cuyo carácter es más tradicional. Analizando eje vertical, en la dimensión positiva, se sitúan establecimientos que en general prestan buenos servicios y trato agradable y en su dimensión negativa, establecimientos que ofrecen menos servicios y un trato menos agradable (véase PROFIT Apartado III.2).

Para más de un segmento

Fase 4: Resultados del análisis multidimensional

4.a) Input del análisis

El análisis multidimensional permite estudiar también qué diferencias existen entre distintos individuos por lo que respecta a la forma en que perciben y posicionan mentalmente un grupo de competidores.

Para desarrollar este apartado, a la información recogida en el apartado anterior, procedente de 100 amas de casa de una localidad 1, se añadirá la correspondiente a otras 100 amas de

TABLA 10.L.4. Distancias promedio entre puntos de venta para el segmento (localidad) 1 y para el segmento (localidad) 2

Localidad 1	*Mercado*	*T.Tradici.*	*Gran Supermercado*	*Supermercado*	*Hipermercado*
Mercado	0	3,04	5,22	5,04	5,37
T.Tradicional	3,04	0	4,83	4,75	5,57
Gran Supermercado	5,22	4,83	0	2,59	3,79
Supermercado	5,04	4,75	2,59	0	3,91
Hipermercado	5,37	5,57	3,79	3,91	0
Localidad 2	*Mercado*	*T.Tradici.*	*Gran Supermercado*	*Supermercado*	*Hipermercado*
Mercado	0	4,27	5,01	5,18	5,17
T.Tradicional	4,27	0	4,33	4,60	5,24
Gran Supermercado	5,01	4,33	0	2,46	2,71
Supermercado	5,18	4,60	2,46	0	2,07
Hipermercado	5,17	5,24	2,71	2,07	0

Nota: escala 1-7, donde 1= nada diferentes y 7= muy diferentes

TABLA 10.L.5. Índices de ajuste

N.º dimensiones	*Promedio RSQ*	*Promedio STRESS*	*Segmento 1*		*Segmento 2*	
			RSQ	*STRESS*	*RSQ*	*STRESS*
1	0,9997	0,00328	1	0,002	1	0,004
2	0,9858	0,5488	0,998	0,21	0,974	0.75

casa de otra localidad 2 en la que operan los mismos competidores ya comentados: mercado, tienda tradicional, un supermercado, un gran supermercado y un hipermercado.

De tal forma, se parte de las percepciones facilitadas por dos grupos, o segmentos, de encuestados de dos localidades diferentes y se trata de ver si existen diferencias entre ellos respecto a como son percibidos los detallistas que operan en ambas localidades.

Para tal fin ha sido utilizado el algoritmo de Análisis de Diferencias Individuales (INDSCAL) incluido en el programa SPSS que, además de posibilitar análisis agregados, permite conocer los pesos o ponderaciones otorgados a cada dimensión por cada uno de los dos segmentos objeto de estudio, localidades en nuestro caso.

Coordenadas en el mapa de los competidores objeto de estudio

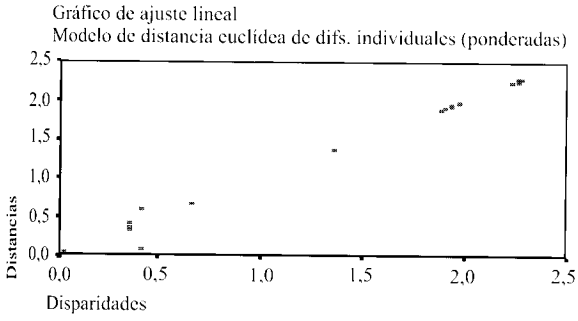
TABLA 10.L.6. Coordenadas de los competidores en la solución bidimensional

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	−1,2203	−1,7624
Tienda tradicional	−1,2154	0,9119
Supermercado	0,7062	0,7224
Gran Supermercado	0,7241	0,5800
Hipermercado	−1,0053	0,4519

TABLA 10.L.7. Pesos de cada segmento sobre cada dimensión

Segmentos	*Dimensión 1*	*Dimensión 2*
Localidad 1	0,9994	0,0238
Localidad 2	0,9544	0,2581
Promedio	0,9548	0,0,336

Gráficos

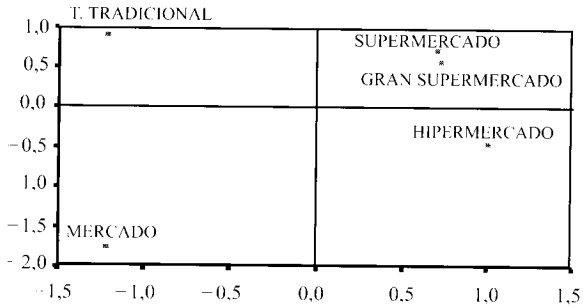


4.b) Resultados del análisis

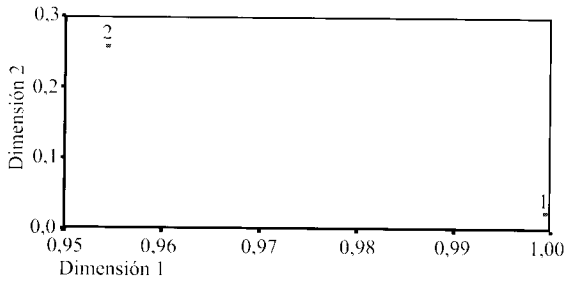
El programa genera diversas salidas, entre las cuales se muestran a continuación: los índices de ajuste (S-Stress, RSQ y Stress),las coordenadas de los establecimientos en el mapa perceptual, diversos gráficos de ajuste y los pesos o ponderaciones de cada segmento (localidad) sobre cada dimensión.

Índices de ajuste

Como se advierte, los índices de ajuste son buenos. El S-Stress ha sido de 0,001 tanto para la solución de dos dimensiones como para la de una[3].



MAPA 2. Posicionamiento competitivo de 5 puntos de venta. Solución agregada.



MAPA 3. Peso derivados de los sujetos. Diferencias individuales.

[3] Como se señaló anteriormente, se podría seguir aumentando el número de dimensiones hasta tres, cuatro, etc. No obstante, dado que en nuestro ejemplo son sólo cinco los competidores objeto de estudio, no tiene mayor sentido continuar haciéndolo.

Fase 5: Interpretación de resultados

En el mapa perceptual 2, se observa como en la dimensión negativa del eje horizontal se sitúan puntos de venta tradicionales y en la dimensión positiva, puntos de venta más modernos. Deteniéndonos en el eje vertical, quedan en la dimensión positiva, los puntos de venta que ofertan buen servicio en general y trato agradable y en el lado opuesto puntos de venta caracterizados por el elevado número de productos vendidos a granel o sin marca, ofreciendo pocos servicios y trato menos agradable (veáse PROFIT Apartado III.2).

En el mapa 3, y en relación con la solución bidimensional, se advierte que el segmento 1 (localidad 1), concede gran importancia al eje horizontal que, como se ha comentado, opone puntos de venta tradicionales frente a fórmulas comerciales más modernas. Es decir, los establecimientos se perciben sobre todo en función del carácter moderno-tradicional que tengan. Por su parte, en el segmento 2 (localidad 2), se concede más importancia al eje vertical. Es decir, los establecimientos se perciben sobre todo en función del servicio que ofertan y el trato prestado. La distancia entre cada segmento (localidad) y el origen de los ejes de coordenadas indica lo bien que está representado en la solución agregada (mapa 2). Mejores ajustes significan mayor distancia del origen. En este caso, se observa que el segmento 1(localidad 1) está peor representado que el segmento 2 (localidad 2), ya que su ordenada se aproxima al origen (0,0238).

II. ANÁLISIS FACTORIAL DE CORRESPONDENCIAS: FASES 4 Y 5

Fase 4: Resultados y ajuste del análisis factorial de correspondencias

4.a) Input del análisis

La información de este análisis recoge la presencia o ausencia de las características en cada establecimiento. Ello requiere contar con una matriz de frecuencias como la recogida en la Tabla 10.L.8. Como se observa, ninguna casilla obtiene más de 100, debido a que se recogieron datos de 100 entrevistadas.

4.b) Resultados

El programa SPSS genera diversas salidas, entre las cuales se muestran: la varianza explicada por cada factor, las coordenadas de los establecimientos y de las propiedades en el mapa, las contribuciones de los puntos (competidores y atributos) a la inercia de cada dimensión y las contribuciones de las dimensiones a la inercia de los puntos (competidores y atributos).

TABLA 10.L.8. Recuento de frecuencias

	Mercado	*T. Tradicional*	*Supermercado*	*Gran Supermercado*	*Hiper*
Tienen todos los productos que necesito	21	16	88	71	55
Está limpio y ordenado	39	48	89	79	64
Tienen marcas conocidas	28	45	89	73	66
Es grande	18	5	65	67	78
Es un establecimiento moderno	16	4	68	73	61
Es fácil encontrar los productos	23	29	81	72	49
Los productos están bien ordenados	30	37	81	82	59
Los precios son altos	20	44	41	*29*	25
Es bueno para comprar productos en oferta	5	4	70	67	50
Son más baratos que otros sitios	7	4	68	59	48
Tienen productos de calidad	37	44	90	79	60
El horario de la tienda me viene bien	28	29	90	58	51
El trato del personal es agradable	42	46	79	72	45
El pago se hace con rapidez	28	28	70	57	39
En los mostradores te atienden rápidamente	32	39	62	45	31
Se puede pagar con tarjeta		2	85	58	56
Está bien localizado	39	30	94	75	50
Dispone de fácil aparcamiento	9	6	75	86	52
Atienden mis problemas con rapidez	30	28	69	59	39
Tienen buena relación calidad precio	33	30	84	69	48
En general ofrecen buen servicio	44	48	90	76	56
Buena valoración global del establecimiento	35	37	88	81	60

Unidad: Frecuencia absoluta y relativa porque la muestra son 100 individuos.

TABLA 10.L.9. Varianza explicada por cada factor

Dimensión	*Valor propio*	*Inercia*	*Varianza explicada*	*Varianza acumulada*
1	0,24788	0,06145	0,771	0,771
2	0,11108	0,01234	0,155	0,925
3	0,06153	0,00379	0,047	0,973
4	0,04655	0,00217	0,027	1,000
Total		0,07974	1,000	1,000

Varianza explicada por cada factor

Reteniendo tan sólo dos factores se puede explicar hasta el 92 por ciento de la variabilidad total de los datos. Resulta particularmente explicativo el factor 1, o primera dimensión.

Coordenadas

Dado que se dispone de una tabla de doble entrada, es posible identificar las coordenadas tanto de los establecimientos como de los atributos.

TABLA 10.L.10. Coordenadas de los competidores en el análisis factorial de correspondencias en la solución bidimensional

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	0,712	–0,006
Tienda tradicional	1,082	–0,295
Eroski	–0,182	0,340
Sabeco	–0,114	–0,173
Pryca	–0,516	–0,543

TABLA 10.L.11. Coordenadas de las propiedades en el análisis factorial de correspondencias en la solución bidimensional

	Dimensión 1	*Dimensión 2*
Tienen todos los productos que necesito	–0,306	0,165
Está limpio y ordenado	0,281	–0,196
Tienen marcas conocidas	0,150	–0,273
Es grande	–0,720	–0,386
Es un establecimiento moderno	–0,668	0,086
Es fácil encontrar los productos	0,003	0,111
Los productos están bien ordenados	0,095	–0,040
Los precios son altos	0,989	–0,551
Es bueno para comprar productos en oferta	–0,783	0,297
Son más baratos que otros sitios	–0,735	0,217
Tienen productos de calidad	0,239	–0,099
El horario de la tienda me viene bien	–0,254	–0,766
El trato del personal es agradable	0,972	0,915
El pago se hace con rapidez	0,214	0,075
En los mostradores te atienden rápidamente	0,651	–0,109
Se puede pagar con tarjeta	–0,923	0,150
Está bien localizado	0,141	0,171
Dispone de fácil aparcamiento	–0,674	0,477
Atienden mis problemas con rapidez	0,232	0,094
Tienen buena relación calidad precio	0,139	0,097
En general ofrecen buen servicio	0,389	–0,099
La valoración global del establecimiento es buena	0,124	–0,030

Contribuciones de los puntos a la inercia de cada dimensión

Para interpretar cada dimensión, es útil analizar en qué medida los distintos puntos (competidores y atributos) contribuyen a su inercia. Por lo que respecta a los competidores, se advierte que la tienda tradicional (0,51) y el mercado (0,209), y en lado opuesto el hipermercado (0,222) contribuyen bastante a la inercia (varianza explicada) de la dimensión 1. En segundo lugar, el gran supermercado (0,28) y en el lado opuesto otra vez el hipermercado (0,55), contribuyen a la inercia de la dimensión dos (Tabla 10.L.12.A).

TABLA 10.L.12.A. Contribuciones de los puntos de venta a la inercia de cada dimensión

Marca	*Dimensiones*	
	1	*2*
1 Mercado	**0,209**	0,000
2 Tienda Tradic.	**0,516**	0,085
3 Gran Supermercado	0,036	**0,280**
4 Supermercado	0,016	0,084
5 Hipermercado	**0,222**	**0,551**
	1,000	1,000

En relación a las propiedades, el factor «precios altos» (0,16) por una parte y el «pago con tarjeta» (0,126) en el lado opuesto, contribuyen a la dimensión 1. Es decir, se oponen establecimientos de precios altos frente establecimientos modernizados que permiten pago con tarjeta. En segundo lugar, por una parte el «trato agradable» (0,32) y en el lado opuesto los «buenos horarios» (0,28), contribuyen a la inercia de la dimensión dos (Tabla 10.L.13.A). Es decir, se oponen establecimientos caracterizados por el trato que prestan frente a establecimientos que destacan por su buen horario.

Contribuciones de las dimensiones a la inercia de los competidores y de las propiedades

Para interpretar los ejes, resulta de gran importancia lo que se conoce como contribuciones de las dimensiones a la inercia de los puntos. En este sentido, se trata de analizar qué puntos (competidores y propiedades) vienen mejor explicados por la dimensión 1 y que puntos (competidores y propiedades) vienen mejor explicados por la dimensión dos. Para tal fin, basta con seleccionar los valores extremos izquierda-derecha para la dimensión 1 o eje horizontal, y los valores extremos arriba-abajo para la dimensión 2 o eje vertical. Estos valores figuran en las Tablas 10.L.12.B y 10.L.13.B y representan la cantidad de variabilidad de un punto explicada por cada dimensión. En el mejor de los casos, la dimensión explica el cien por cien de la variabilidad del punto.

TABLA 10.L.12.B. Contribución de las dimensiones a la inercia de los puntos de venta competidores (OTR)

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	0,835	0,000
Tienda tradicional	0,939	0,031
Supermercdo	0,336	0,526
Gran Supermercado	0,288	0,298
Hipermercado	0,666	0,331

La dimensión primera, o eje horizontal, explica en su parte negativa el comportamiento del hipermercado (OTR = 66 por ciento) y en su parte positiva el de la tienda tradicional (OTR = 93 por ciento) y el mercado (OTR = 83 por ciento). La dimensión segunda, o eje vertical, explica en su parte negativa el comportamiento del hipermercado (OTR = 33 por ciento) y en su parte positiva el del supermercado (OTR = 52 por ciento) y el gran supermercado (OTR = 29 por ciento).

La dimensión primera, o eje horizontal, explica en su parte negativa propiedades como «es moderno» (OTR = 91 por ciento), «bueno para comprar ofertas» (OTR = 90 por ciento), «son más baratos» (OTR = 94 por ciento) y en su parte positiva aspectos como «tienen productos de calidad» (OTR = 91 por ciento), «atienden rápidamente» (OTR = 97 por ciento) y «ofrecen buen servicio» (OTR = 94 por ciento). La dimensión segunda, o eje vertical, explica en su parte negativa propiedades como «tienen buenos horarios»

TABLA 10.L.13.A. Contribuciones de las columnas a la inercia de cada dimensión

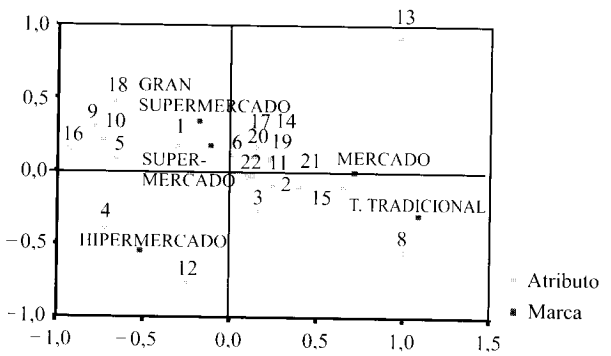
Atributo	*Dimensiones*	
	1	*2*
1	0,017	0,011
2	0,018	0,020
3	0,005	0,037
4	0,088	0,057
5	0,072	0,003
6	0,000	0,005
7	0,002	0,001
8	0,114	0,079
9	0,088	0,028
10	0,074	0,014
11	0,013	0,005
12	0,014	**0,289**
13	**0,165**	**0,327**
14	0,007	0,002
15	0,065	0,004
16	**0,126**	0,007
17	0,004	0,014
18	0,076	0,085
19	0,009	0,003
20	0,004	0,004
21	0,035	0,005
22	0,003	0,000
	1.000	1.000

(OTR = 76 por ciento) y en su parte positiva aspectos como «es fácil encontrar los productos» (OTR = 39 por ciento), «el trato es agradable» (OTR = 28 por ciento) y «tiene buena localización» (OTR = 18 por ciento).

TABLA 10.L.13.B. Contribución de las dimensiones a la inercia de las propiedades (OTR)

	Dimensión 1	*Dimensión 2*
Tienen todos los productos que necesito	0,805	0,105
Está limpio y ordenado	0,793	0,173
Tienen marcas conocidas	0,281	0,416
Es grande	0,803	0,103
Es un establecimiento moderno	**0,918**	0,007
Es fácil encontrar los productos	0,001	**0,398**
Los productos están bien ordenados	0,319	0,026
Los precios son altos	0,781	0,108
Es bueno para comprar productos en oferta	**0,908**	0,059
Son más baratos que otros sitios	**0,944**	0,037
Tienen productos de calidad	**0,910**	0,070
El horario de la tienda me viene bien	0,188	**0,766**
El trato del personal es agradable	0,715	**0,284**
El pago se hace con rapidez	0,843	0,046
En los mostradores te atienden rápidamente	**0,974**	0,012
Se puede pagar con tarjeta	0,856	0,010
Está bien localizado	0,286	**0,188**
Dispone de fácil aparcamiento	0,762	0,171
Atienden mis problemas con rapidez	0,779	0,057
Tienen buena relación calidad precio	0,554	0,120
En general ofrecen buen servicio	**0,941**	0,028
La valoración global del establecimiento es buena	0,767	0,020

Gráficos



Tienen todos los productos que necesito	1
Está limpio y ordenado	2
Tienen marcas conocidas	3
Es grande	4
Es un establecimiento moderno	5
Es fácil encontrar los productos	6
Los productos están bien ordenados	7
Los precios son altos	8
Es bueno para comprar productos en oferta	9
Son más baratos que otros sitios	10
Tienen productos de calidad	11

El horario de la tienda me viene bien	12
El trato del personal es agradable	13
El pago se hace con rapidez	14
En los mostradores te atienden rápidamente	15
Se puede pagar con tarjeta	16
Está bien localizado	17
Dispone de fácil aparcamiento	18
Atienden mis problemas con rapidez	19
Tienen buena relación calidad precio	20
En general ofrecen buen servicio	21
La valoración global del establecimiento es buena	22

MAPA 4. Posicionamiento competitivo de 5 puntos de venta y de las propiedades que los definen.

Fase 5: Interpretación de resultados

Tras obtener un mapa perceptual en el que se recogen tanto los atributos como los establecimientos objeto de estudio, puede procederse a la interpretación de su significado. En primer lugar, las distancias entre objetos deben interpretarse de igual modo que en el análisis multidimensional, si bien en este caso la incorporación de las características supone un elementos adicional de interpretación.

En segundo lugar y, atendiendo al eje horizontal puede decirse que el hipermercado se percibe como «moderno», «barato» y «con buenas ofertas» frente al mercado y la tienda tradicional que se perciben como puntos de venta más orientados a la «calidad» y al «servicio» y en consecuencia más caros. Atendiendo al eje vertical, el hipermercado se percibe como un punto de venta con «buenos horarios» frente al supermercado y al gran supermercado que se preocupan más por la «localización», por ofrecer productos «fáciles de encontrar en el establecimiento» y un «trato agradable».

Por último, la varianza explicada por cada uno de los ejes ofrece información adicional para interpretar la cantidad de información explicada por cada factor. Así, el eje horizontal explica el 77 por ciento de la varianza, mientras que el eje vertical, tan sólo el 15 por ciento.

III. PROGRAMAS ESPECIALES: FASES 4 Y 5

III.1. Análisis MDPREF

Fase 4: Resultados y ajuste del análisis multidimensional de preferencias MDPREF

4.a) Input del análisis

Si se desea representar en que medida en cada uno de los dos segmentos (localidades) se prefieren los distintos establecimientos, se dispone de programas específicos como el MDPREF para ello. Su mayor utilidad radica en permitir la representación simultánea de competidores e individuos/segmentos sobre un mismo espacio perceptual. Si no se deseara analizar la población de forma segmentada, esto es, si se quisiera estudiar la estructura de preferencias promedio de toda la población, bastaría un único punto para representarla en el mapa.

El input que sería la estructura de preferencias promedio hacia los distintos establecimientos, podría ser el que se indica en la tabla 10.L.14.

TABLA 10.L.14. Preferencias promedio de cada segmento

	Mercado	*T. Tradici.*	*Gran Supermercado*	*Supermercado*	*Hipermercado*
Segmento 1	2,46	3,64	3,71	**1,86**	4,32
Segmento 2	3,65	4,17	3,37	**2,26**	4,17

Nota: 1 = el primero en la jerarquía de preferencias, 5 = el último.

No resulta difícil comprobar que el establecimiento más preferido es el supermercado tanto en la localidad 1 (segmento 1), como en la localidad 2 (segmento 2).

4.b) Resultados del análisis

Coordenadas

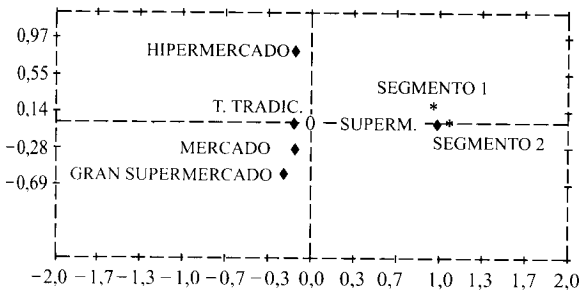
TABLA 10.L.15. Coordenadas de los competidores en la solución bidimensional

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	–0,2012	–0,2383
Tienda tradicional	–0,1925	0,0509
Supermercado	0,8932	0,0045
Gran Supermercado	–0,2523	–0,5882
Hipermercado	–0,2472	0,7711

TABLA 10.L.16. Coordenadas de los segmentos en la solución bidimensional

Segmentos	*Dimensión 1*	*Dimensión 2*
1	0,09933	1,1159
2	1,0000	–0,0076

Gráficos



MAPA 5. Posicionamiento competitivo de 5 puntos de venta y de las preferencias de los dos segmentos (localidades).

Fase 5: Interpretación de resutados

La estructura de preferencias de ambos segmentos (localidades) es muy parecida. No en vano, su correlación es de 0,9924. Ambos prefieren el supermercado respecto al resto de establecimientos. El segmento 2, parece más inclinado hacia el mercado y el segmento 1 hacia la tienda tradicional y el mercado. Ambos se encuentran alejados del hipermercado.

III.2. Análisis PROFIT de ajuste de propiedades

Fase 4: Resultados y ajuste del análisis de ajuste de propiedades

4.a) Input del análisis

Para interpretar el significado de los ejes de coordenadas del mapa generado con el análisis multidimensional (punto I), resulta posible incorporar *a posteriori* una serie de propiedades. Simplemente se necesita conocer la puntuación promedio que cada establecimiento ha recibido sobre las propiedades que se quieren proyectar sobre un espacio perceptual previamente obtenido (véase punto I). De este modo, se puede completar el análisis multidimensional de escalas para conseguir, como en el Análisis Factorial de Correspondencias, que además de representar competidores, se representen las propiedades o atributos sobre las que éstos compiten.

En esta matriz se advierte que el gran supermercado y el supermercado son los que obtienen puntuaciones promedio más elevadas para todas las propiedades analizadas. El Hipermercado, aunque a unos niveles más bajos, también destaca en algunas propiedades mucho más que el mercado y la tienda tradicional. Concretamente destaca frente a ambos en las propiedades 1, 2, 3, 4, 5 y 6. El mercado y la tienda tradicional, no destacan frente al resto en ninguna propiedad. El hipermercado tampoco destaca en las propiedades 7 y 8.

TABLA 10.L.17. Coordenadas de los competidores en la solución bidimensional del MDS

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	1,6052	–0,5341
Tienda tradicional	1,5162	0,3842
Supermercado	–0,9619	0,5222
Gran Supermercado	–0,8067	0,5006
Hipermercado	–1,3528	–0,8730

TABLA 10.L.18. Puntuaciones promedio de cada competidor sobre cada una de las 8 propiedades objeto de estudio

	Marcas 1	*Moderno 2*	*Fácil 3*	*Ofertas 4*	*Calidad 5*	*Horario 6*	*Trato 7*	*Servicio 8*
Mercado	2,8	1,6	2,3	0,5	3,7	2,8	4,2	4,4
Tienda tradicional	4,5	0,4	2,9	0,4	4,4	2,9	4,6	4,8
Gran Supermercado	**7,3**	**7,3**	**7,2**	**6,7**	**6,6**	**5,8**	**7,2**	**7,6**
Supermercado	**8,9**	**6,8**	**8,1**	**7,0**	**9,0**	**9,0**	**7,9**	**9,0**
Hipermercado	**6,6**	**6,1**	**4,9**	**5,0**	**6,0**	**5,1**	4,5	5,6

4.b) Resultados de análisis

El programa PROFIT permite obtener los resultados siguientes: proyecciones de los vectores de las 8 propiedades objeto de estudio, índices de ajuste y coordenadas de las propiedades sobre el mapa perceptual que se obtuvo en un análisis previo.

Proyecciones

TABLA 10.L.19. Proyecciones de los vectores ajustados de las 8 propiedades objeto de estudio

	Marcas 1	*Moderno 2*	*Fácil 3*	*Ofertas 4*	*Calidad 5*	*Horario 6*	*Trato 7*	*Servicio 8*
Mercado	–1,4739	–1,6938	–1,3828	–1,6272	–1,4140	–1,4446	–1,0474	–1,2267
Tienda tradicional	–0,7236	–1,3248	–0,5810	–1,0251	–0,6281	–0,6759	–0,1438	–0,3647
Gran Supermercado	**1,0320**	**1,0762**	**0,9900**	**1,0890**	**1,0048**	**1,0189**	**0,8158**	**0,9118**
Supermercadoi	**0,9124**	**0,9219**	**0,8812**	**0,9492**	**0,8923**	**0,9028**	**0,7435**	**0,8204**
Hipermercado	**0,2532**	**1,0206**	**0,0927**	**0,6141**	**0,1450**	**0,1989**	–0,3681	–1,1407

En la Tabla 10.L.19 se recogen las proyecciones sobre cada propiedad de cada uno de los ocho establecimientos.

Solo el gran supermercado y el supermercado tienen proyecciones positivas para todas las propiedades analizadas. Ello obedece a que, como se ha indicado, ambos establecimientos son los únicos que destacan en términos de promedio en todas las propiedades. El hipermercado también tiene buena parte de sus proyecciones positivas. Concretamente las relativas a los atributos 1, 2, 3, 4, 5 y 6. La razón estriba en que el hipermercado también destaca en estas propiedades (Tabla 10.L.18). Ahora bien, como destaca en estas propiedades menos que el supermercado y el gran supermercado, los valores de sus proyecciones son inferiores. El mercado y la tienda tradicional tienen sus proyecciones negativas sobre todas las propiedades ya que en promedio no destacan en ninguna de ellas. El hipermercado también tiene proyecciones negativas para las propiedades 7 y 8 ya que no destaca en ninguna de ambas.

TABLA 10.L.20. Ajuste o correlación entre las puntuaciones originales de las 8 propiedades y las proyecciones de los vectores que las representan

	RHO	*Coeficiente de correlación*	
1. Tienen marcas conocidas	0,9616	R = 0,962	RSQ = 0,925
2. Es un establecimiento moderno	0,9656	R = 0,966	RSQ = 0,932
3. Es fácil encontrar los productos	0,9621	R = 0,962	RSQ = 0,926
4. Bueno para comprar ofertas	0,9759	R = 0,976	RSQ = 0,952
5. Tiene productos de calidad	0,8848	R = 0,885	RSQ = 0,783
6. El horario me va bien	0,8848	R = 0,845	RSQ = 0,714
7. El trato del personal es agradable	0,9245	R = 0,924	RSQ = 0,855
8. En general ofrecen buen servicio	0,9159	R = 0,916	RSQ = 0,839

Índices de ajuste

En la Tabla 10.L.19 se recogen distintos indicadores que miden cómo las propiedades aparecen representadas en el mapa perceptual. Concretamente, el RSQ (R × R), mide la correlación entre las puntuaciones originales de las propiedades (Tabla 10.L.18) y sus proyecciones (Tabla 10.L.19). Equivale al porcentaje de varianza de los datos que puede ser explicada por el modelo. Lo ideal es que se pueda explicar el cien por cien de la variabilidad de cada propiedad. Es decir, que el índice se aproxime a la unidad. Rho es otro índice de ajuste que también permite medir como de bien se representan las puntuaciones originales de cada propiedad sobre el mapa obtenido, analizando la correlación entre las proyecciones de los cinco establecimientos sobre el vector ajustado «X» y sus corespondientes valores originales para esa propiedad «X». Valores próximos a la unidad, indican mejor ajuste.

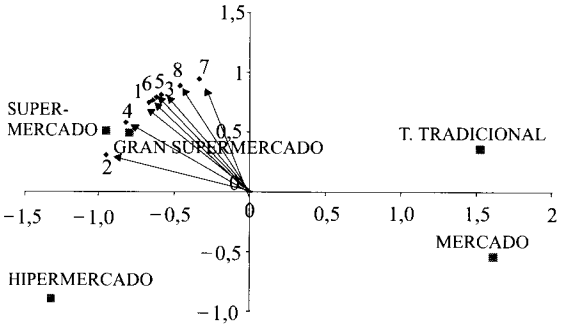
Se observa que las proyecciones de los 5 establecimientos sobre la propiedad 4, aproximan realmente bien la puntuaciones promedio originales de los establecimiento sobre esta propiedad. Para el resto de propiedades también existe una alta correspondencia entre los valores originales (Tabla 10.L.18) y las proyecciones (Tabla 10.L.19).

Coordenadas de las propiedades

TABLA 10.L.21. Coordenadas de las propiedades incorporadas a la solución bidimensional

	Dimensión 1	*Dimensión 2*
1. Tienen marcas conocidas	0,6670	0,7451
2. Es un establecimiento moderno	0,9518	0,3066
3. Es fácil encontrar los productos	0,5890	0,8082
4. Bueno para comprar ofertas	0,8213	0,5705
5. Tiene productos de calidad	0,6150	0,7885
6. El horario me va bien	0,6412	0,7674
7. El trato del personal es agradable	0,3348	0,9423
8. En general ofrecen buen servicio	0,4658	0,8849

Gráficos



MAPA 6. Posicionamiento competitivo de 5 puntos de venta y de 8 propiedades incorporadas a posteriori sobre un espacio perceptual.

Fase 5: Interpretación de resultados

Como se recoge en el mapa 1 tan sólo el supermercado, el gran supermercado y el hipermercado se proyectan sobre las propiedades estudiadas positivamente. Concretamente el supermercado y el gran supermercado, se proyectan con valores mejores que el hipermercado. Los tres establecimientos destacan especialmente por tener marcas conocidas (1), ser modernos (2) y ofrecer buenos precios (4). Por su parte, el mercado y la tienda tradicional no se perciben ni modernos, ni económicos.

El hipermercado, dado que apenas destaca en las propiedades 7 y 8 frente al resto de establecimientos, se proyecta en el plano sobre ambas propiedades de forma mucho peor de lo que lo hace sobre las otras seis propiedades. Es decir, ni presta trato agradable (7), ni ofrece buenos servicios (8). La tienda tradicional y el mercado no destacan en ninguna de las propiedades analizadas, ya que al proyectarse en el plano sobre ellas, lo hacen de forma negativa. Es decir, no tienen marcas conocidas, ni son modernos, ni ofrecen buenos precios. Además, al igual que el hipermercado, ni prestan trato agradable ni ofrecen buen servicio en general, especialmente el mercado.

III.3. Análisis PREFMAP

Fase 4: Resultados y ajuste del análisis de ajuste de preferencias

4.a) Input del análisis

De la misma manera que se han incorporado las *propiedades a posteriori*, sobre un mapa perceptual previamente generado, también se pueden incorporar a posteriori las preferencias de los segmentos objeto de estudio.

Es decir, a menudo puede resultar interesante extender el análisis de percepciones al de preferencias. Con el método de incorporación externa de preferencias (PREFMAP) que se desarrolla a continuación, se toma como punto partida un mapa perceptual generado previamente y sobre el se ajustan las preferencias de distintos sujetos. Es decir, los dos inputs del programa son, en primer lugar, las coordenadas de cada establecimiento sobre un mapa perceptual y en segundo lugar, las preferencias de los distintos sujetos (localidades) objeto de análisis hacia tales establecimientos.

TABLA 10.L.22. Coordenadas de los competidores en la solución bidimensional del MDS

Competidores	*Dimensión 1*	*Dimensión 2*
Mercado	1,6052	−0,5341
Tienda tradicional	1,5162	0,3842
Supermercado	−0,9619	0,5222
Gran Supermercado	−0,8067	0,5006
Hipermercado	−1,3528	−0,8730

TABLA 10.L.23. Preferencias promedio de cada segmento

	Mercado	*T. Tradici.*	*Gran supermercado*	*Supermercado*	*Hipermercado*
Segmento 1	**2,46**	**3,64**	3,71	**1,86**	4,32
Segmento 2	3,65	4,17	3,37	**2,26**	4,17

Nota: 1 = el primero en la jerarquía de preferencias, 5 = el último.

No resulta difícil comprobar que el establecimiento más preferido es el supermercado tanto en la localidad 1 (segmento 1), como en la localidad 2 (segmento 2). El hipermercado ocupa los peores lugares en el ranking de preferencias promedio en ambas localidades.

4.b) Resultados de análisis

Coordenadas

Este programa opera en cuatro fases. Los resultados de la primera fase, que es la más general, son los que a continuación se indican. En esta fase, se permite que cada sujeto siga su propia orientación y que conceda sus propios pesos a las dimensiones.

Se observa que el supermercado, el establecimiento más preferido, es el más próximo (menos distante) respecto al punto ideal. El hipermercado, que es el menos preferido, es el más alejado del punto ideal.

TABLA 10.L.24. Coordenadas de los segmentos en la solución bidimensional

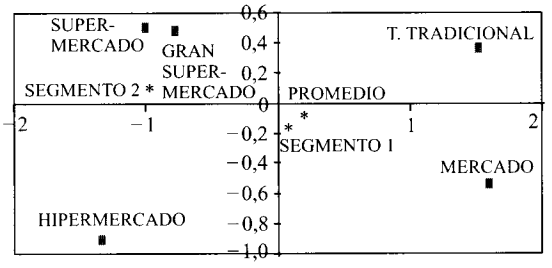
Segmento (localidad 1)	0,09947	–0,20438
Segmento (localidad2)	–0,97761	0,05566
Segmento promedio	0,13866	–0,17581

TABLA 10.L.25. Distancias cuadradas entre cada competidor y cada segmento

	Mercado	*T. Tradici.*	*Gran Supermercado*	*Supermercado*	*Hipermercado*
Segmento 1	**5,8497**	**6,1026**	7,0436	**6,0830**	6,9948
Segmento 2	**–0,0864**	**0,1013**	0,3012	**0,2880**	0,4109
Segmento promedio	**2,8637**	3,0648	3,6490	**3,1681**	4,1011

Nota: 1 = el primero en la jerarquía de preferencias, 5 = el último.

Gráficos



MAPA 7. Posicionamiento competitivo de 5 puntos de venta y de las preferencias de dos segmentos y del segmento promedio.

Fase 5: Interpretación de resultados

En este mapa, además de representarse los establecimientos desde el punto de vista de cómo son percibidos, se introducen las preferencias de los individuos (localidades) hacia los mismos. Es decir, se combinan percepciones y preferencias en un mismo espacio perceptual.

El segmento 1 parece más inclinado hacia la tienda tradicional y el mercado, y el segmento 2 hacia el gran supermercado. Ambos se encuentran alejados del hipermercado.

Aunque el análisis de preferencias trata de medir en qué medida distintos individuos (segmentos) prefieren distintos establecimientos, si se deseara realizar el análisis de preferencias tan sólo para un sólo individuo (segmento), simplemente se incorporarían sus preferencias sobre el mapa ya obtenido siguiendo el mismo procedimiento de actuación.

SECCIÓN

4

TÉCNICAS AVANZADAS Y EMERGENTES

Valoración general

Esta sección proporciona una introducción simple y concisa para algunas de las técnicas más innovadoras que están surgiendo actualmente en el análisis multivariante. En muchas ocasiones, la adopción de una nueva técnica se ve retrasada por culpa de una «mística» adoptada por los expertos que no quieren o no pueden pasar su conocimiento a los demás. No obstante, se ofrece a los investigadores de hoy día una gama de técnicas nuevas y desafiantes, por lo que tienen a su disposición un abanico más amplio de métodos y pueden extender su capacidad para abordar problemas completamente nuevos. Por tanto, hoy día existe la necesidad real de la introducción de estas técnicas que no solamente proporcionan una comprensión general de los procedimientos, sino también el conocimiento de cuándo se pueden aplicar y a qué problemas se ajustan de forma adecuada. No hemos podido cubrir todo lo que es nuevo en el análisis multivariante. En su lugar, hemos seleccionado un avance fundamental en el tratamiento de las relaciones de dependencia múltiple (los modelos de ecuaciones estructurales) y varias técnicas que han surgido del dominio aplicado. Esperemos que este equilibrio proporcione una cobertura adecuada de los progresos que tienen lugar tanto en la comunidad académica como en la de los investigadores aplicados.

Capítulos de la Sección 4

La Sección 4 contiene dos capítulos. El Capítulo 11 aborda los modelos de ecuaciones estructurales (SEM), un procedimiento para acoplar el error de medición directamente a la estimación de una serie de relaciones de dependencia. Los modelos SEM proporcionan dos ventajas claras que no se hallaban en las técnicas multivariantes abordadas hasta ahora: (1) la capacidad de incorporar directamente el error de medición en el proceso de estimación, y (2) la estimación simultánea de varias relaciones de dependencia interrelacionada. Los modelos de ecuaciones estructurales

han sido más ampliamente aceptados por la comunidad académica, pero han logrado una reducida utilización en otras áreas debido en parte a la tremenda «curva de aprendizaje» que se percibe como necesaria para su aplicación. No queremos subestimar el esfuerzo involucrado, pero ningún investigador debería evitar el SEM simplemente por esta razón porque los principios del análisis factorial y la regresión múltiple forman los métodos subyacentes que se utilizan en el SEM. El Capítulo 12 introduce una gama de aplicaciones emergentes que han sido desarrolladas principalmente en el ámbito de la aplicación. Procedimientos como el almacenaje y la búsqueda de datos se han visto estimulados por las grandes cantidades de información que se recogen en las organizaciones actualmente y la necesidad de análisis oportunos y concisos. Los investigadores de hoy se enfrentan a enormes problemas con el tratamiento de grandes bases de datos que pueden contener cientos de miles o millones de casos mientras intentan abordar las cuestiones de investigación que requieren un componente de exploración. ¿Cómo podemos identificar las soluciones potenciales de entre las posibles miles alternativas? Además, ¿cómo podemos apalancar estas grandes bases de datos y «explotarlas» extrayendo las «ideas fundamentales» en la información que podrían proporcionar? Estas son cuestiones que se abordan en el Capítulo 12.

CAPÍTULO 11

Modelos de ecuaciones estructurales

OBJETIVOS DE APRENDIZAJE

- Entender el papel de las relaciones causales en los análisis estadísticos.
- Representar una serie de relaciones causales en un diagrama de secuencias.
- Trasladar un diagrama de secuencias a un conjunto de ecuaciones para su estimación.
- Apreciar el papel y la influencia de las propiedades de medición de las variables sobre los resultados de los análisis estadísticos.
- Evaluar los resultados del análisis de los modelos de ecuaciones estructurales como apoyo de las relaciones propuestas y las posibles áreas de mejora de los resultados.
- Aplicar las técnicas de los modelos de ecuaciones estructurales a problemas tales como el análisis de factorial confirmatorio, análisis de secuencias y estimación de ecuaciones simultáneas.

Uno de los objetivos primordiales de las técnicas multivariantes es aumentar la capacidad explicativa del investigador y su eficacia estadística. La regresión múltiple, el análisis factorial, el análisis multivariante de la varianza, el análisis discriminante y otras técnicas discutidas en los capítulos precedentes proporcionan al investigador un conjunto de poderosas herramientas a la hora de tratar con un amplio abanico de cuestiones prácticas y teóricas. Pero todas ellas tienen una limitación común: cada técnica puede examinar sólo una relación al mismo tiempo. Incluso las técnicas que tienen en cuenta varias variables dependientes, como el análisis multivariante de la varianza y el análisis canónico, siguen representando sólo una única relación entre las variables dependientes e independientes.

Demasiado a menudo, sin embargo, el investigador se enfrenta a un conjunto de cuestiones interrelacionadas. Por ejemplo, ¿cuántas variables determinan la imagen de una tienda? ¿Cómo se combina esa imagen con otras variables para afectar a las decisiones de compra y a la satisfacción con la tienda? Y finalmente, ¿cómo se convierte la satisfacción con la tienda en la lealtad a largo plazo? Esta serie de asuntos tienen importancia tanto teórica como práctica. Ninguna de estas técnicas nos permite tratar con todas estas cuestiones a la vez mediante una técnica exhaustiva. Por esta razón, examinamos la técnica del **modelo de ecuaciones estructurales (SEM)**, una extensión de varias técnicas multivariantes que ya hemos estudiado, entre ellas la regresión múltiple y el análisis factorial.

Como se describe brevemente en el Capítulo 1, el modelo de ecuaciones estructurales examina simultáneamente una serie de relaciones de dependencia. Es particularmente útil cuando una variable dependiente se convierte en variable independiente en ulteriores relaciones de dependencia. Este conjunto de relaciones, cada una con variables dependientes e independientes, es la base del SEM. La formulación básica del SEM en forma ecuacional es

$$
\begin{aligned}
Y_1 &= X_{11} + X_{12} + X_{13} + \cdots + X_{1n} \\
Y_2 &= X_{21} + X_{22} + X_{23} + \cdots + X_{2n} \\
Y_m &= X_{m1} + X_{m2} + X_{m3} + \cdots + X_{mn} \\
&\text{(métrica)} \qquad \text{(métrica, no métrica)}
\end{aligned}
$$

El modelo de ecuaciones estructurales se ha utilizado en casi todos los campos de estudio concebibles, incluyendo la educación, el marketing, la psicología, la sociología, la gestión, la contrastación y medida, la salud, la demografía, el comportamiento organizacional, la biología e incluso la genética. Las razones de su atractivo para áreas tan diversas son dobles: (1) proporciona un método directo de tratar con múltiples relaciones simultáneamente a la vez que se da eficacia estadística, y (2) su capacidad para evaluar las relaciones exhaustivamente y proporcionar una transición desde el **análisis exploratorio** al **confirmatorio**. Esta transición corresponde a los mayores esfuerzos en todos los campos de estudio hacia el desarrollo de perspectivas más sistemáticas y holísticas de los problemas. Tales esfuerzos exigen la capacidad de contrastar una serie de relaciones que constituyen un modelo a gran escala, un conjunto de principios fundamentales o una teoría entera. Estas son tareas para las que el modelo de ecuaciones estructurales se ajusta perfectamente.

¿Qué es el modelo de ecuaciones estructurales?

El modelo de ecuaciones estructurales (SEM) abarca una familia entera de modelos conocidos con muchos nombres, entre ellos análisis de la estructura de la covarianza, análisis de variable latente, análisis de factor confirmatorio y a menudo simplemente análisis LISREL (el nombre de uno de los programas de *software* más populares). Producto de una evolución de la modelización multiecuacional desarrollada principalmente en la econometría y fusionada con los principios de medición de la psicología y la sociología, el SEM se ha convertido en una herramienta integral tanto en la investigación académica como en la práctica [7, 11, 12, 20, 25, 39, 41, 51, 61, 63, 71, 80, 96]. SEM puede utilizarse también como medio de estimación de otros modelos multivariantes, incluyendo la regresión, componentes principales [37], correlación canónica [40] e incluso MANOVA [9].

Como puede esperarse de una técnica de uso tan extendido y con tantas variantes y aplicaciones, muchos investigadores no están seguros de lo que constituye el modelo de ecuaciones estructurales. Sin embargo, todas las técnicas SEM se distinguen por dos características: (1) estimación de relaciones de dependencia múltiples y cruzadas, y (2) la capacidad de representar **conceptos no observados** en estas relaciones y tener en cuenta el error de medida en el proceso de estimación.

Tratamiento de relaciones múltiples de dependencia cruzada

La diferencia más obvia entre SEM y otras técnicas de relaciones multivariantes es el uso de relaciones distintas para cada conjunto de variables dependientes. En los términos más sencillos, SEM estima una serie de ecuaciones de regresión múltiples distintas pero interrelacionadas mediante la especificación del **modelo estructural** utilizado por el programa estadístico. En primer lugar, el investigador utiliza la teoría, la experiencia previa y los objetivos de investigación para diferenciar qué variables independientes predicen cada variable dependiente. En nuestro ejemplo de los capítulos anteriores, queríamos en primer lugar predecir la imagen del establecimiento. A continuación queríamos utilizar la imagen del establecimiento para predecir la satisfacción y utilizar a la vez las dos variables para predecir la fidelidad al establecimiento. Por tanto, algunas variables dependientes se convierten en variables independientes en relaciones ulteriores, dando lugar a la naturaleza interdependiente del modelo estructural. Además, muchas de las mismas variables afectan a cada una de las variables dependientes, pero con efectos distintos. El modelo estructural expresa estas relaciones entre variables dependientes e independientes, incluso cuando una variable dependiente se convierte en variable independiente en otras relaciones.

Las relaciones propuestas se trasladan a continuación a series de ecuaciones estructurales (parecidas a las ecuaciones de regresión) para cada variable dependiente. Esta característica sitúa a SEM en un lugar aparte de las técnicas discutidas previamente que trataban variables de dependencia múltiples —análisis multivariante de la varianza y de la correlación canónica— en la que sólo se permite una *única* relación entre las variables dependientes e independientes.

Incorporación de variables que no se miden directamente

La estimación de relaciones múltiples interrelacionadas no es el único elemento de la modelización de ecuaciones estructurales. SEM también tiene la habilidad de incorporar **variables latentes** al análisis. Una variable latente es un concepto supuesto y no observado que sólo puede ser aproximado mediante variables medibles u observables. Las variables observadas, que recogemos a través de varios métodos de obtención de datos (es decir, estudios de mercado, experimentos, observación) se conocen como **variables manifiestas**. Entonces, ¿por qué querríamos utilizar una variable latente no medida en lugar de datos exactos (variables manifiestas) que ofrecen los encuestados? Aunque esto pueda parecer que no tiene sentido o un enfoque de «caja negra», tiene justificaciones tanto prácticas como teóricas, al mejorar la estimación estadística, mejorar la representación de conceptos teóricos y tener en cuenta el error de medida.

Mejora de la estimación estadística

La teoría estadística nos dice que un coeficiente de regresión está compuesto en realidad de dos elementos: el «verdadero» o coeficiente estructural entre la variable dependiente e independiente y la **fiabilidad** de la variable predictor. La fiabilidad es el grado en el que la variable independiente está «libre de error» [19]. En todas las técnicas multivariantes expuestas hasta ahora, hemos asumido que no existía error en nuestras variables. Pero sabemos tanto desde la perspectiva práctica como teórica que no podemos medir perfectamente un concepto y que siempre hay algún grado de **error de medida**. Por ejemplo, cuando preguntamos acerca de algo tan inmediato como la renta de las economías domésticas, sabemos que algunas personas contestarán incorrectamente, tanto por exceso como por defecto o no lo sabrán con precisión. Las respuestas ofrecidas tienen cierto error de medida y por tanto afectan al «verdadero» coeficiente estructural [87].

El impacto del error de medida (y la correspondiente disminución de la fiabilidad) puede observarse de una expresión del coeficiente de regresión como la siguiente:

$$\beta_{y \cdot x} = \beta_s \times \rho_x$$

donde $\beta_{v,x}$ es el coeficiente de regresión observado, β_s es el «verdadero» coeficiente estructural y ρ_x es la fiabilidad de la variable predictor. A menos que la fiabilidad sea del 100 por cien, la correlación observada siempre infravalora la relación «correcta». Dado que todas las relaciones de dependencia se basan en la correlación observada (y resultante del coeficiente de regresión) entre las variables, esperaremos «fortalecer» las correlaciones utilizando un modelo de dependencia para hacer estimaciones más precisas de los coeficientes estructurales teniendo en cuenta en primer lugar la correlación atribuible a cualquiera de los problemas de medida.

Representación de conceptos teóricos

El error de medición no sólo está provocado por respuestas inadecuadas sino que se produce también cuando utilizamos conceptos teóricos o más abstractos, tales como la actitud hacia un producto o las motivaciones del comportamiento. Con conceptos tales como estos, el investigador intenta diseñar las mejores cuestiones para medir el concepto [82]. Los encuestados también pueden no estar seguros de cómo responder o de cómo interpretar las cuestiones de forma diferente a la que pretende el investigador. Ambas situaciones pueden dar lugar al error de medida. Pero si sabemos la magnitud del problema, podemos incorporar la fiabilidad en la estimación estadística y mejorar nuestro modelo de dependencia.

Especificación del error de medida

¿Cómo tenemos en cuenta el error de medida? SEM proporciona el **modelo de medida**, que especifica las reglas de correspondencia entre las variables latentes y manifiestas. El modelo de medida permite al investigador utilizar una o más variables para un único concepto dependiente o independiente y a continuación estimar (o especificar) la fiabilidad. Por ejemplo, la variable dependiente puede ser un concepto representado por un conjunto de cuestiones (como la escala aditiva introducida en el Capítulo 3). En el modelo de medida el investigador puede evaluar la contribución de cada ítem de la escala, así como incorporar el grado en que la escala mide el concepto (su fiabilidad) en la estimación de las relaciones entre las variables dependientes e independientes. Este procedimiento es similar al desarrollo por el análisis factorial de los ítems de la escala y utilizar las puntuaciones del factor en la regresión. Estas similitudes y los detalles específicos se discuten en una sección posterior de este capítulo.

Un ejemplo sencillo de SEM

La modelización de ecuaciones estructurales proporciona al investigador la capacidad de acomodar relaciones de dependencia cruzadas múltiples en un único modelo. Su analogía más próxima es la regresión múltiple, que puede estimar una única relación (ecuación). Pero SEM puede estimar varias ecuaciones a la vez que pueden estar interrelacionadas, lo que quiere decir que la variable dependiente en una ecuación puede ser una variable independiente en otra(s) ecuación(es). Esto permite al investigador modelizar relaciones complejas que no serían posibles con ninguna otra de las técnicas multivariantes que hemos discutido en este texto.

El siguiente ejemplo ilustra cómo opera SEM con relaciones múltiples. Hay que observar, sin embargo, que no representa uno de los otros puntos fuertes del SEM: la capacidad para emplear múltiples medidas de un concepto de forma similar al análisis factorial. El Capítulo 3 introduce el concepto de formación de escalas para representar conceptos o utilizar puntuaciones de factores como sustitutos de conjuntos de variables. SEM puede desarrollar un procedimiento similar cuando el investigador especifica el conjunto de variables que representan cada concepto. Para simplificar, asumimos que cada concepto del siguiente ejemplo se mide por una única variable. Trataremos los beneficios de los indicadores múltiples posteriormente dentro de este capítulo, pero ahora nos centraremos sólo en los principios básicos de la estimación de relaciones múltiples.

La cuestión de investigación

Los ejecutivos de HATCO han esperado mucho tiempo para aumentar la productividad y el mantenimiento de los empleados en la empresa a través de una mejor comprensión de sus motivaciones y actitudes hacia HATCO. Con este fin, el departamento de personal ha identificado tres actitudes que creen que son las más importantes: satisfacción con el trabajo, el compromiso con la organización y la probabilidad de movilidad de trabajadores. A continuación desarrollaron relaciones que vinculan cada actitud con sus antecedentes en una relación de dependencia:

Variable dependiente	*Variables independientes*
Satisfacción con el trabajo	= Compañerismo + Entorno de trabajo
Lealtad de la organización	= Satisfacción con el trabajo + Nivel de sueldo
Probabilidad de movilidad de los trabajadores	= Satisfacción con el trabajo + Lealtad a la organización

Identificación del modelo de ecuación estructural para el análisis de secuencias

Cuando el departamento de personal encarga su trabajo al departamento de investigación, los investigadores se dan cuenta de que la regresión múltiple no estimaría las relaciones y que en su lugar hay que utilizar SEM. Para representar mejor las relaciones de interdependencia, crean una representación gráfica de las relaciones (véase Figura 11.1) que denominan **diagrama de secuencias**. Las flechas directas señalan el impacto de las variables independientes sobre las dependientes y las flechas curvadas señalan la correlación entre variables, igual que la multicolinealidad en la regresión múltiple.

¿Cuál era el propósito de desarrollar el diagrama de secuencias? Los diagramas de secuencias son los fundamentos del **análisis de relaciones**, el procedimiento de estimación empírica de la fortaleza de cada relación (secuencia) representada en el diagrama de secuencias. El análisis de relaciones calcula la fortaleza de las relaciones utilizando sólo una matriz de correlación o covarianza

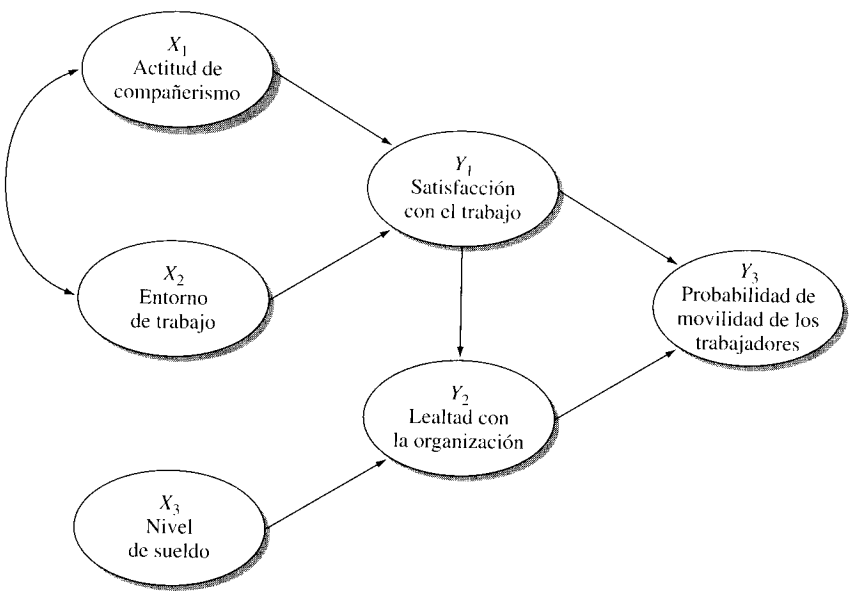


FIGURA 11.1. Diagrama de secuencias de las relaciones entre las actitudes de los empleados.

como input. La correlación simple (bivariante) entre cualesquiera dos variables puede ser representada como la suma de las secuencias combinadas que conectan estos puntos. Una secuencia combinada es una secuencia de flechas que sigue tres normas:

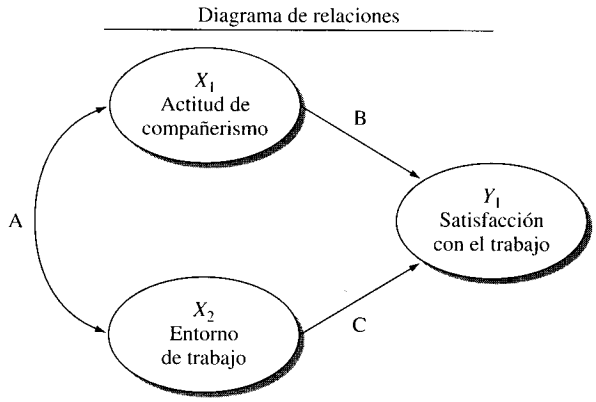
1. Después de ir hacia delante con una flecha, no puede volver hacia atrás; pero la secuencia puede volver cuantas veces sea necesario antes de seguir hacia delante.
2. La secuencia no puede ir por la misma variable más de una vez.
3. La secuencia puede incluir sólo una flecha curvada (par de variables correlacionadas).

Una aplicación del análisis de secuencias

Estas pueden parecer reglas muy complicadas, pero en realidad son muy simples. Podemos ilustrar cómo operan aplicándolas a partes pequeñas de nuestro diagrama de secuencias. En la Figura 11.2 hemos escogido una relación de dependencia simple para la satisfacción en el trabajo que es parte de un modelo más amplio. Las dos variables independientes de actitud de compañerismo (X_1) y entorno de trabajo (X_2) están correlacionadas y predicen la variable dependiente de satisfacción en el trabajo (Y_1). Esto puede representarse de una forma tan simple como esta:

$$Y_1 = b_1X_1 + b_2X_2$$

El análisis de relaciones nos permite utilizar correlaciones sencillas entre constructos para estimar las relaciones causales representadas por los coeficientes b_1 y b_2. Para facilitar las secuencias, las relaciones causales se denominan A, B y C. La relación causal A es X_1 correlacionada con X_2, la relación B es el efecto de X_1 prediciendo Y_1 y la relación C muestra el efecto de X_2 prediciendo Y_1.



Correlaciones bivariantes

	X_1	X_2	Y_1
X_1	1,0		
X_2	0,50	1,0	
Y_1	0,60	0,70	1,0

Correlaciones como relaciones compuestas

$\text{Corr}_{X_1X_2} = A$
$\text{Corr}_{X_1Y_1} = B + AC$
$\text{Corr}_{X_2Y_1} = C + AB$

Resolución de coeficientes estructurales

0,50 = A
0,60 = B + AC
0,70 = C + AB

Sustituyendo A = 0,50
0,60 = B + 0,50C
0,70 = C + 0,50B

Resolviendo para B y C
B = 0,33
C = 0,53

FIGURA 11.2. Cálculo de los coeficientes estructurales con el análisis de relaciones.

El análisis de relaciones utiliza las correlaciones simples tal y como se muestra en la Figura 11.2 para estimar las relaciones causales utilizando las tres reglas dadas anteriormente. Por ejemplo, la correlación de X_1 e Y_1 ($r_{x_1y_1} = 0{,}60$) puede representarse como dos relaciones causales: B y AC. El símbolo B representa la relación directa de X_1 con Y_1 y la otra relación (una secuencia compuesta) sigue la línea curvada de X_1 a X_2 y después a Y_1. Asimismo, la correlación de X_2 e Y_1 puede demostrarse que está compuesta por dos relaciones causales: C y AB. Finalmente, la correlación de X_1 y X_2 es igual a A. Esta relación forma tres ecuaciones.

$$r_{x_1x_2} = A$$
$$r_{x_1y_1} = B + AC$$
$$r_{x_2y_1} = C + AB$$

Sabemos que A es igual a 0,50, así que podemos sustituir este valor en las otras ecuaciones. Resolviendo estas dos ecuaciones, conseguimos los valores de B (b_1) = 0,33 y C (b_2) = 0,53. Los cálculos efectivos se muestran en la Figura 11.2. Este enfoque permite al análisis de relaciones resolver cualquier relación causal basada sólo en las correlaciones entre los constructos y el modelo causal especificado. Como puede verse a partir de este sencillo ejemplo, si cambiamos el modelo de secuencias de alguna forma, las relaciones causales cambiarán también. Tal cambio proporciona la base de la modificación del modelo para conseguir un mejor ajuste, si está teóricamente justificado.

Con estas reglas simples, el modelo más amplio se puede modelizar ahora simultáneamente utilizando correlaciones o covarianzas como datos de entrada. Notaremos que cuando utilizamos un modelo superior, podemos resolver las ecuaciones interrelacionadas. Por ejemplo, resolver la ecuación

Probabilidad de movilidad del empleado = satisfacción en el trabajo + lealtad a la organización

se hace con las mismas reglas que en nuestro ejemplo anterior. Por tanto, las variables dependientes en una relación pueden ser fácilmente variables independientes en otra relación. No importa lo grande que sea el diagrama de secuencias o cuántas relaciones se incluyan, el análisis de secuencias ofrece una forma de analizar el conjunto de relaciones.

Resumen

El análisis de relaciones puede extenderse a cualquier sistema de relaciones. Con estos procedimientos, se pueden estimar todas las relaciones de cualquier diagrama de secuencias para cuantificar los efectos entre las variables dependientes y las independientes, incluso si están interrelacionadas. El investigador tiene así una técnica para analizar un conjunto de relaciones de forma simultánea. Hemos presentado sólo una simple ilustración del proceso; a los lectores interesados se les anima a examinar otros tratamientos de este asunto [33, 38, 68, 99].

La función de la teoría en los modelos de ecuaciones estructurales

A lo largo de la exposición del SEM, nos referiremos a la necesidad de una justificación teórica para la especificación de las relaciones de dependencia, modificaciones a las relaciones propuestas y muchos otros aspectos de la estimación de un modelo. La «teoría» ofrece la causa de casi todos los aspectos del SEM. A efectos de este capítulo, la **teoría** puede definirse como un conjunto sistemático de relaciones que ofrecen una explicación exhaustiva y consistente de un fenómeno. A partir de esta definición, podemos ver que teoría no es el dominio exclusivo de los académicos sino que puede estar basada en la experiencia y en la práctica obtenida por la observación del com-

portamiento del mundo real. La teoría es a menudo un objetivo prioritario de la investigación académica, pero la investigación práctica puede desarrollar o proponer un conjunto de relaciones que son tan complejas e interrelacionadas como cualquier teoría de base académica. Por tanto, tanto los investigadores académicos como los de la empresa pueden beneficiarse de unas herramientas analíticas extraordinarias proporcionadas por SEM.

Desde una perspectiva práctica, un enfoque teórico de SEM es una necesidad porque la técnica debe ser casi completamente especificada por el investigador. Mientras que con otras técnicas multivariantes el investigador puede haber sido capaz de especificar un modelo básico y permitir que valores por defecto de los programas estadísticos «completen» los resultados de estimación restantes, SEM no tiene ninguna de estas características. Aunque el proceso de siete fases que discutimos hace estas decisiones directas, debe definirse explícitamente cada componente de los modelos y medidas estructurales. Además, cualesquiera modificaciones del modelo deben hacerse a través de acciones específicas dictadas por el investigador. La necesidad de un modelo «teórico» para guiar el proceso de estimación se hace especialmente crítica cuando se hacen modificaciones del modelo. Dada la flexibilidad de SEM, las oportunidades de «sobreajustar» el modelo o desarrollar un modelo con escasa generalización son muy elevadas. Por tanto, cuando resaltamos la necesidad de la justificación teórica, nuestro objetivo es que el investigador reconozca que SEM es un método confirmatorio, guiado más por la teoría que por los resultados empíricos.

Desarrollo de una estrategia de modelización

Uno de los conceptos más importantes que un investigador debe aprender en relación con las técnicas multivariantes es que no existe un único modo «correcto» de aplicarlas. En lugar de eso, el investigador debe formular los objetivos de la investigación y aplicar la técnica de la forma más apropiada para conseguir los objetivos deseados. En algunos casos, las relaciones están especificadas estrictamente y el objetivo es una confirmación de la relación. Otras veces, las relaciones están vagamente reconocidas y el objetivo es el descubrimiento de las relaciones. En cada caso límite y en los casos intermedios, el investigador debe formular el uso de la técnica de acuerdo con los objetivos de investigación.

La aplicación de SEM sigue este mismo principio. Su flexibilidad ofrece al investigador una poderosa herramienta de análisis apropiada para muchos objetivos de investigación. Pero el investigador debe definir estos objetivos como líneas maestras de la estrategia de modelización. El uso del término *estrategia* se refiere a un plan de acción orientado a la consecución de un objetivo específico. En el caso de SEM, el resultado último es siempre la evaluación de una serie de relaciones. Sin embargo, esto se puede conseguir de muchas maneras. Para nuestros propósitos, definimos tres estrategias distintas en la aplicación de SEM: estrategia de modelización confirmatoria, estrategia de modelos rivales y estrategia de desarrollo del modelo.

Estrategia de modelización confirmatoria

La aplicación más directa de los modelos de ecuaciones estructurales es una **estrategia de modelización confirmatoria**, donde el investigador especifica un modelo aislado y SEM se utiliza para evaluar su significación estadística. Aquí el investigador está diciendo, «tanto si funciona como si no». Aunque pueda parecer que es la aplicación más rigurosa, en realidad no es el test más riguroso del modelo propuesto. La investigación ha mostrado incluso que las técnicas desarrolladas para evaluar los modelos de ecuaciones estructurales tienen un *sesgo confirmatorio*, que tiende a confirmar que el modelo se ajusta a los datos. Por tanto, si el modelo propuesto tiene un ajuste aceptable por cualquiera de los criterios aplicados, el investigador no ha «probado» el modelo propuesto, sino que sólo ha confirmado que es uno de los varios modelos posibles aceptables. Varios modelos pueden tener ajustes igualmente aceptables. Por tanto, el test más riguroso se consigue comparando modelos rivales.

Estrategia de modelos rivales

Obtener un nivel de ajuste aceptable tanto para el modelo conjunto como para los modelos de medida o estructural no asegura al investigador que se haya encontrado el «mejor» modelo. Varios modelos alternativos pueden ofrecer iguales o incluso mejores ajustes. Como medio de evaluar el modelo estimado con modelos alternativos, las comparaciones de modelos conjuntos pueden llevarse a cabo en una **estrategia de modelos rivales**. El test más fuerte de un modelo propuesto es identificar y contrastar los modelos rivales que representan las verdaderas y distintas relaciones hipotéticas estructurales. Cuando se comparan estos modelos, el investigador se acerca mucho al contraste de «teorías» alternativas, que es un contraste mucho más fuerte que una ligera modificación de una única «teoría».

¿Cómo genera el investigador este conjunto de modelos alternativos? Un posible origen de modelo rivales son las formulaciones alternativas de la teoría subyacente. Por ejemplo, en una formulación la confianza puede preceder a la lealtad mientras que en la otra la lealtad precede a la confianza. Esto podría constituir la base de dos modelos rivales. Los **modelos equivalentes** proporcionan una segunda perspectiva sobre el desarrollo de un conjunto de modelos rivales. Se ha mostrado que para cualquier modelo de ecuaciones estructurales, existe al menos otro modelo con el mismo número de parámetros y el mismo nivel de ajuste del modelo que varía en las relaciones representadas. Esto implica que ningún modelo es único en el nivel de ajuste conseguido, y que para cualquier modelo con un ajuste aceptable existe un número indeterminado de modelos alternativos con el mismo nivel de ajuste. Se han definido una serie de normas para identificar modelos equivalentes para cualquier modelo estructural [69]. Como norma general, cuanto más complejo sea el modelo, más modelos equivalentes existen. Una tercera aproximación, el programa TETRAD, es un método empírico que está tomando auge y que examina sistemáticamente un modelo estructural e identifica relaciones adicionales que están apoyadas por los datos [50]. Con la matriz de datos de entrada y la especificación del modelo hecha por el investigador, el programa examina las pautas de las relaciones (tetrads) y aisla aquellas relaciones que pudieran estar empíricamente fundamentadas. El programa TETRAD no estima los parámetros, sino que identifica las relaciones que se van a incluir en el modelo original para formar los modelos rivales. El programa funciona mejor empezando con un modelo simple y añadiendo relaciones. Se suele alegar que este modelo es ateórico y muy mecanicista o de «caja negra», pero ofrece a los investigadores intuiciones para sus modelos que no se podrían haber obtenido de otra forma.

Un ejemplo habitual de la estrategia de modelos rivales es el proceso de evaluación factorial de invarianza, la igualdad de modelos de factor entre grupos. Existe un procedimiento establecido para evaluar el grado de invarianza, comenzando por los modelos más vagamente restringidos y a partir de esos modelos añadir paulatinamente restricciones adicionales hasta que se contraste el modelo más restrictivo [75]. Las restricciones se añaden para representar invarianzas a lo largo de grupos, ponderaciones e incluso intercorrelaciones de factores. Esto también es un ejemplo de un enfoque de **modelos anidados**, en el que el número de constructos e indicadores permanece constante, pero el número de relaciones estimadas cambia. Aunque los modelos rivales son normalmente modelos anidados, también pueden no estar anidados (difieren en el número de constructos o indicadores), con lo que se exigen medidas especializadas de ajuste para comparar entre los modelos.

Estrategia de desarrollo del modelo

La **estrategia de desarrollo del modelo** difiere de las dos anteriores estrategias en que aunque se propone un modelo, el propósito del esfuerzo de modelización es mejorarlo a través de modificaciones de los modelos de medida y/o estructurales. En muchas aplicaciones, la teoría sólo puede ofrecer un punto de partida para el desarrollo de un modelo con justificación teórica que pueda ser apoyado empíricamente. Por tanto, el investigador ha de emplear SEM no sólo para contrastar el modelo empíricamente sino también para obtener perspectivas acerca de su reespecificación. Debe tomarse, sin embargo, alguna precaución. El investigador tiene que ser cuidadoso no em-

pleando esta estrategia en la medida en que el modelo final tenga un ajuste aceptable pero que no pueda ser generalizada a otras muestras o poblaciones. Además, la reespecificación del modelo debe hacerse siempre con apoyo teórico en lugar de justificación empírica.

Pasos en la modelización de ecuaciones estructurales

El verdadero valor de SEM proviene de los beneficios de utilizar los modelos estructurales y de medida simultáneamente, jugando cada uno de ellos distintos papeles en el análisis conjunto. Para asegurar que ambos modelos están correctamente especificados y que los resultados son válidos, discutiremos ahora un proceso en siete etapas (véase Figura 11.3, que muestra los pasos 1-3, y Figura 11.6, que muestra los pasos 4-7). Este proceso difiere de la construcción de modelos en seis pasos introducidos en el Capítulo 1 y utilizados en nuestras discusiones acerca de otros métodos multivariantes. La introducción de este proceso distinto para SEM no invalida el enfoque de construcción de modelos para otras técnicas multivariantes, sino que sólo acentúa la singularidad de SEM.

Los siete pasos de la modelización de ecuaciones estructurales son: (1) desarrollar un modelo fundamentado teóricamente, (2) construir un diagrama de secuencias de relaciones causales, (3) convertir el diagrama de secuencias en un conjunto de modelos y relaciones estructurales, (4) elegir el tipo de matriz de entrada y estimar el modelo propuesto, (5) evaluar la identificación del modelo estructural, (6) evaluar los criterios de calidad del ajuste, y (7) interpretar y modificar el modelo si está teóricamente justificado.

Primer paso: Desarrollo de un modelo basado en la teoría

La modelización de ecuaciones estructurales se basa en **relaciones causales**, en las que el cambio en una variable se supone que produce un cambio en otra variable [57]. Encontramos este tipo de enunciado cuando definimos una relación de dependencia, tal y como se encuentra en el análisis de regresión. Las relaciones causales pueden tomar muchas formas y significados, desde la estricta causación encontrada en los procesos físicos, tales como una reacción química, a las relaciones menos definidas encontradas en la investigación del comportamiento, tales como las «causas» del éxito educativo o las «razones» por las cuales compramos un producto y no otro. La fuerza y convicción con que el investigador puede asumir la **causación** entre dos variables no descansa en los métodos analíticos escogidos sino en la justificación teórica ofrecida por los análisis. Los «requisitos» para hacer valer la causación tienen profundas raíces en varios enfoques de la filosofía de la ciencia [8, 27, 60]. Hay un acuerdo general con al menos cuatro criterios establecidos para realizar afirmaciones causales: (1) asociaciones suficientes entre dos variables, (2) antecedentes temporales de la causa frente al efecto, (3) falta de alternativas a las variables causales, y (4) una base teórica para la relación. Aunque en muchos casos todos los criterios establecidos para la realización de afirmaciones causales no se cumplen estrictamente, probablemente puedan realizarse afirmaciones causales si las relaciones están basadas en una teoría racional. Pero avisamos a cualquier investigador contra la suposición de que las técnicas discutidas en este capítulo ofrezcan por sí mismas un medio de «probar» la causación sin tener una perspectiva teórica como guía. Utilizar estas técnicas de forma «exploratoria» es inútil y equivoca al investigador más que ofrecerle resultados apropiados.

El error más crítico en el desarrollo de modelos de fundamentación teórica es la omisión de una o más variables predictivas claves, un problema conocido como **error de especificación**. La consecuencia de omitir una variable significativa consiste en sesgar la evaluación de la importancia de otras variables [72]. Por ejemplo, si suponemos que dos variables (*a* y *b*) fueran predictores de *c*, e incluyéramos tanto *a* como *b* en nuestro análisis, haríamos la evaluación correcta de su impor-

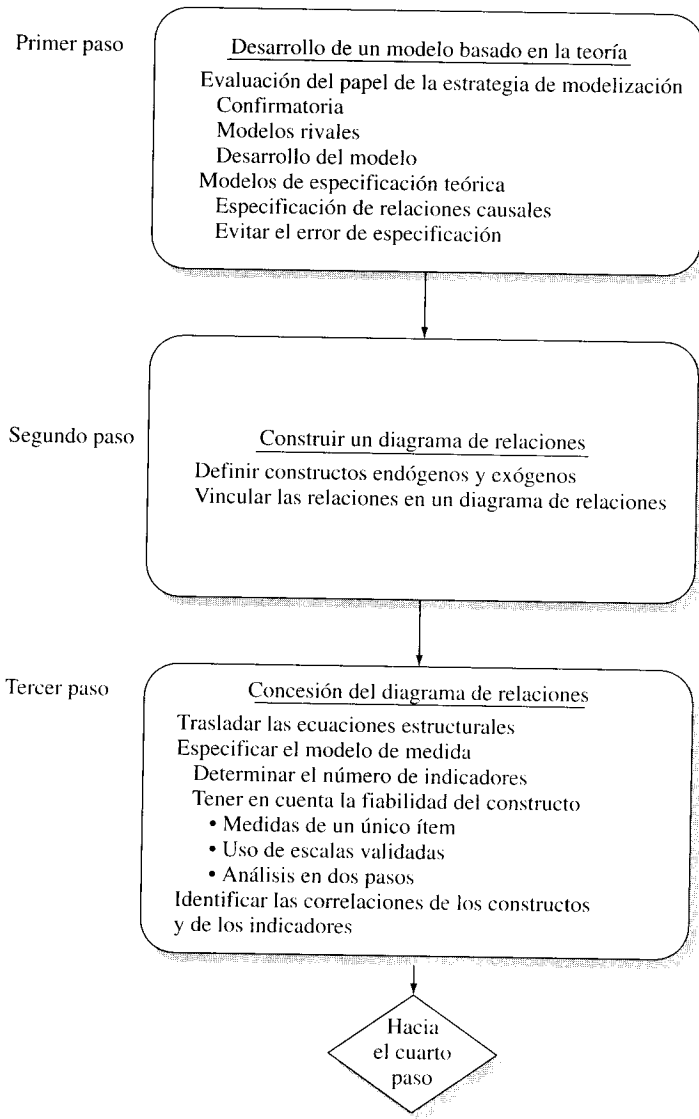


FIGURA 11.3. Pasos 1-3 en el proceso de siete fases para la modelización de ecuaciones estructurales.

tancia relativa como se muestra por sus coeficientes estimados. Pero si dejamos a la variable *b* fuera de nuestro análisis, el coeficiente para *a* sería diferente. Esta diferencia, o sesgo, es el resultado del coeficiente para la variable *a* que refleja no sólo su efecto sobre *c* sino el efecto que comparte también con *b*. Este efecto compartido, sin embargo, está controlado cuando ambas variables están incluidas en el análisis. Animamos a los lectores que todavía no hayan comprendido el impacto de las variables omitidas a revisar el material del Capítulo 4 que se centra en los errores de especificación y sus efectos.

El deseo de incluir todas las variables debe compensarse con las limitaciones prácticas de SEM. Aunque no existe un límite teórico sobre el número de variables del modelo, los problemas prácticos aparecen incluso antes de llegar a los límites de la mayoría de los problemas informáticos. Muy a menudo, las interpretaciones de los resultados, particularmente la significación estadística, se hace bastante difícil a medida que el número de conceptos aumenta (más de 20 conceptos). El investigador nunca debería omitir un concepto solamente porque el número de variables se haga grande, pero también debería de reconocer los beneficios de los modelos teóricos concisos y parsimoniosos.

Segundo paso: Construcción de un diagrama de secuencias de relaciones causales

Hasta el momento, hemos expresado las relaciones causales sólo en términos de ecuaciones. Pero existe otro método de representar estas relaciones denominadas **diagramas de secuencias**, que son especialmente útiles en el diseño de una serie de relaciones causales. Un diagrama de secuencias es más que una simple representación visual de las relaciones porque permite al investigador presentar no sólo las relaciones predictivas entre constructos (es decir, las relaciones variable dependiente-independiente), sino también relaciones asociativas (correlaciones) entre los constructos e incluso entre los indicadores. Discutiremos las implicaciones de cada tipo de relación en secciones ulteriores de este capítulo. Aquí tratamos el diagrama de secuencias, que presenta un método conciso para expresar cada uno de estos tipos de relaciones.

Elementos de un diagrama de secuencias

Antes de examinar los diagramas de secuencias, debemos definir dos elementos básicos utilizados en su construcción. La primera es el concepto de un **constructo**, que es un concepto teórico que actúa como una «piedra angular» utilizada para definir relaciones. Un constructo puede representar un concepto tan simple como edad, renta o género o tan complejo como estatus socioeconómico, conocimiento, preferencia o actitud. El investigador define diagramas de secuencia en términos de constructos y a continuación encuentra variables para medir cada constructo. Por ejemplo, podemos preguntar la edad a alguien y utilizar esto como medida del constructo edad. De la misma manera, podemos preguntar una serie de cuestiones acerca de las opiniones de una persona y utilizar esto como una medida de actitud. Ambas series de cuestiones ofrecen valores numéricos para los constructos. Podemos evaluar las cuestiones para la cantidad de error medido que poseen e incluir esto en el proceso de estimación. A partir de este momento, utilizamos el término «constructo» para representar un concepto particular, no importa cómo se mida. Un constructo se representa normalmente en un diagrama de secuencias mediante un óvalo.

El segundo elemento básico es la flecha, utilizada para representar relaciones específicas entre constructos. Una flecha directa indica una relación causal directa de un constructo a otro. Una flecha curvada (o una línea sin cabezas de flechas) entre constructos indica simplemente una correlación entre constructos. Finalmente, una flecha directa con dos cabezas indica una relación **recíproca** o **no recursiva** entre constructos. Sólo con estos tres tipos de flechas, se pueden representar todas las relaciones de los modelos estructurales.

Ejemplos de diagramas de secuencias

La Figura 11.4 muestra tres ejemplos de relaciones representadas por diagramas de secuencias, junto con las correspondientes ecuaciones. La Figura 11.4a muestra un modelo simple de dos constructos. Tanto X_1 como X_2 son constructos predictores para Y_1, y la flecha curvada entre X_1 y X_2 muestra los efectos de intercorrelaciones (multicolinealidad) sobre la predicción. Podemos mostrar esta relación con una simple ecuación, al igual que hicimos en nuestra discusión de regresión múltiple.

En Figura 11.4b, añadimos un segundo constructo dependiente, Y_2. Ahora, además del modelo y la ecuación mostrada en la Figura 11.4a, añadimos una segunda ecuación mostrando la relación entre X_2 e Y_1 con Y_2. Aquí podemos ver en primer lugar el papel único jugado por SEM cuando más de una relación «comparte» constructos. Queremos saber los efectos de X_1 sobre Y_1, los efectos de X_2 sobre Y_1 y simultáneamente los efectos de X_2 e Y_1 sobre Y_2. Si no las estimamos de forma consistente, no estaríamos seguros de representar sus efectos «verdaderos y aislados». Por ejemplo, se necesita esa técnica para mostrar los efectos de X_2 tanto sobre Y_1 como sobre Y_2.

Las relaciones se complican más en la Figura 11.4c, con tres constructos dependientes, cada uno relacionado con el resto y con los constructos independientes. Existen incluso relaciones re-

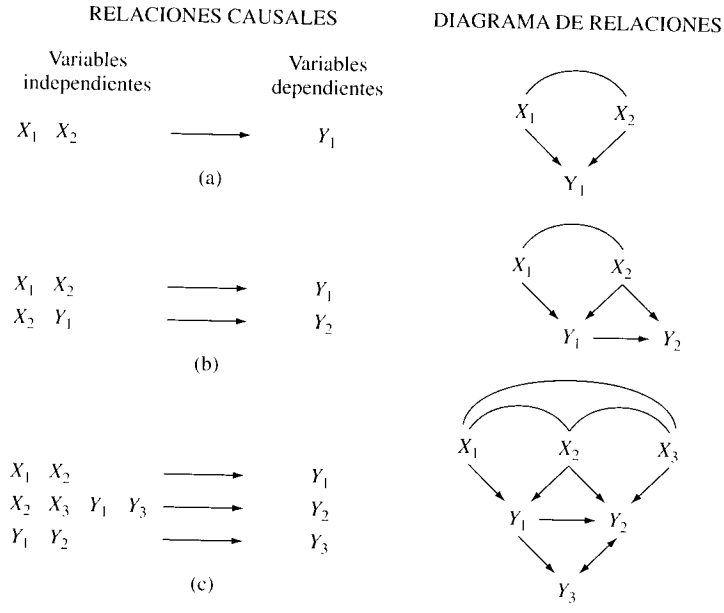


FIGURA 11.4. Representación de las relaciones causales a través de diagramas de secuencias.

cíprocas entre Y_2 e Y_3. Esto se muestra en las ecuaciones por la aparición de Y_2 como predictor de Y_3, e Y_3 como predictor de Y_2. No es posible expresar todas las relaciones ni en la Figura 11.4b ni en la 11.4c en una misma ecuación. Se requieren ecuaciones aisladas para cada constructo dependiente. La necesidad de un método que pueda estimar todas las ecuaciones simultáneamente sólo la cumple SEM.

Terminología básica

Antes de empezar, introduzcamos la terminología necesaria para distinguir entre los constructos de nuestros diagramas de secuencias. Todos los constructos de un diagrama de secuencias pueden clasificarse en dos clases: exógenos y endógenos. Los **constructos exógenos**, también conocidos como variables origen o variables independientes, no están «causados» o son explicados por ninguna de las variables del modelo; esto es, no existen flechas apuntando a estos constructos. En nuestros ejemplos de la Figura 11.4, X_1, X_2 y X_3 eran constructos exógenos.

En las Figuras 11.4a, 11.4b y 11.4c, los constructos Y son **endógenos**, lo que significa que se predicen mediante uno o más constructos. Los constructos endógenos pueden predecir otros constructos endógenos (aquí es donde vemos las interrelaciones que apuntan a la necesidad de los modelos estructurales), pero un constructo exógeno puede estar causalmente relacionado sólo con los constructos endógenos. Por tanto, la distinción entre exógenos y endógenos los determina sólo el investigador, al igual que se hizo al decidir sobre cuáles son variables dependientes e independientes en la regresión.

Supuestos de un diagrama de secuencias

Dos hipótesis subyacen en los diagramas de secuencias. En primer lugar, todas las relaciones causales están indicadas. Por tanto, la teoría es la base para la inclusión o la omisión de cualquier relación. Es tan importante justificar por qué *no* existe una relación causal entre dos constructos como justificar la existencia de otra relación. Pero es importante recordar que el objetivo es modelizar las relaciones entre constructos con el número más reducido de secuencias causales o correlaciones entre constructos que pueden estar teóricamente justificados. El segundo supuesto se refiere a

la naturaleza de las relaciones causales que se supone que son lineales. Al igual que se encontró en otras técnicas multivariantes, las relaciones no lineales no pueden ser estimadas directamente en la modelización de ecuaciones estructurales, pero modelos estructurales modificados pueden aproximar relaciones no lineales [54, 70].

Tercer paso: Conversión de un diagrama de secuencias en un conjunto de ecuaciones estructurales y especificación del modelo de medida

Después de desarrollar el modelo teórico y de representarlo en un diagrama de secuencias, el investigador está preparado para especificar el modelo en términos más formales. Esto se hace a través de una serie de ecuaciones que definen (1) las ecuaciones estructurales que vinculan los constructos, (2) el modelo de medida que especifica qué variables miden qué constructos y (3) una serie de matrices que indican cualquier correlación supuesta entre constructos o variables. El objetivo es vincular definiciones operacionales de los constructos con la teoría para llegar al contraste empírico apropiado. En esta sección se discute cada tipo de ecuación o matriz. En las discusiones, todas las ecuaciones se especifican en terminología general. Los lectores interesados en terminología matemática específica, como la notación utilizada en LISREL, pueden verla en el Apéndice 11A.

Modelo estructural

Trasladar un diagrama de secuencias a una serie de ecuaciones estructurales es un procedimiento directo. En primer lugar, cada constructo endógeno (cualquier constructo con una o más flechas directas hacia él) es la variable independiente en una ecuación separada.

Entonces las variables predictor son todos constructos en los principios o «colas», de las flechas que apuntan a la variable endógena. Es así de simple.

La Figura 11.5 ilustra este proceso de transición para cada uno de los diagramas de secuencias de la Figura 11.4. Como vemos, cada variable endógena (Y_j) puede ser prevista tanto por variable(s) exógena(s) (X_j) como por otras variable(s) endógena(s). Para cada efecto supuesto, estimamos un coeficiente estructural (b_{jm}). Además, dado que sabemos que tendremos errores de predicción, al igual que en la regresión múltiple, incluimos un término de error (ε_i) para cada ecuación. El término de error representa la suma de los efectos debidos a un error de especificación y error aleatorio de medida. No es posible separar estas dos fuentes de error excepto en situaciones especiales.

	VARIABLES ENDÓGENAS	=	VARIABLES EXÓGENAS	+	VARIABLES ENDÓGENAS	+	ERROR
	Y_1		X_1 X_2 X_3		Y_1 Y_2 Y_3		ε_i
Diagrama de relaciones							
Figura 11.4(a)	Y_1	=	$b_1X_1 + b_2X_2$			+	ε_1
Figura 11.4(b)	Y_1	=	$b_1X_1 + b_2X_2$			+	ε_1
	Y_2	=	b_3X_2	+	b_4Y_1	+	ε_2
Figura 11.4(c)	Y_1	=	$b_1X_1 + b_2X_2$			+	ε_1
	Y_2	=	$b_3X_2 + b_4X_3$	+	$b_5Y_1 + b_6Y_3$	+	ε_2
	Y_3	=		+	$b_7Y_1 + b_8Y_2$	+	ε_3

FIGURA 11.5. Transición de un diagrama de secuencias a ecuaciones estructurales.

Modelo de medida

Hasta ahora nos hemos referido al término de medida en términos generales, pero ahora debemos definirlo en términos específicos. Discutiremos no sólo los procedimientos básicos de especificación de un modelo de medida, sino también los supuestos referentes al número de indicadores por constructo y el proceso de especificación de la fiabilidad del constructo en lugar de en su estimación. Pero antes de hacerlo, revisemos los fundamentos del análisis factorial (Capítulo 3), que son muy parecidos a los del modelo de medida.

Correspondencia con el análisis factorial En el análisis factorial, cada variable individual se «explicaba» por su ponderación en cada factor. El objetivo era representar lo mejor posible todas las variables en un número reducido de factores, es decir, los factores referidos a «dimensiones subyacentes» de los datos, que después tendremos que interpretar y clasificar. El análisis factorial, tal y como se discutió en el Capítulo 3, a menudo se clasifica como una técnica exploratoria porque no existen restricciones sobre las cargas de las variables. Cada variable tiene una carga sobre cada factor. Representamos estas relaciones matemáticamente en la Tabla 11.1.

El valor de cada factor (puntuación del factor) se calcula mediante las cargas sobre cada variable (es decir, factor 1 $= L_{11}V_1 + L_{21}V_2 + L_{31}V_3 + L_{41}V_4 + L_{51}V_5$, donde V_1 hasta V_5 son los valores efectivos de cada variable).También el valor predictor para cada variable se calcula mediante las cargas de la variable para cada factor. Sin embargo, cada variable tiene una carga factorial; por tanto, cada factor es siempre una composición de todas las variables, aunque sus cargas varíen en magnitud. Por consiguiente, un factor es en realidad un constructo latente, definido por las cargas de todas las variables.

TABLA 11.1. Comparación de las cargas del análisis factorial e indicadores de las ponderaciones del modelo de medida

	Análisis factorial: Cargas factoriales			*Modelo de medida: Indicadores de las ponderaciones en cada constructo*		
Variable	*Factor* 1	*Factor* 2	*Factor* 3	*Constructo A*	*Constructo B*	*Constructo C*
V_1	L_{11}	L_{12}	L_{13}	L_1		
V_2	L_{21}	L_{22}	L_{23}	L_2		
V_3	L_{31}	L_{32}	L_{33}		L_3	
V_4	L_{41}	L_{42}	L_{43}		L_4	
V_5	L_{51}	L_{52}	L_{53}			L_5

Especificación del modelo de medida Para especificar el modelo de medida, hacemos la transición desde el análisis factorial, en el que el investigador no tiene el control sobre qué variables describen cada factor, a un modo **confirmatorio**, en el que el investigador especifica qué variables definen cada constructo (factor). Las variables observadas que obtenemos de los encuestados se denominan **indicadores** en el modelo de medida, porque los utilizamos para medir o «indicar», los constructos latentes (factores). Supongamos en nuestro ejemplo que V_1 y V_2 se supone que son indicadores del constructo A, V_3 y V_4 son indicadores del constructo B, y V_5 es un indicador único del constructo C. El modelo de medida quedaría expresado como se indica en la Tabla 11.1.

¿Cómo y por qué difiere esta configuración de las cargas del análisis factorial examinadas previamente? La diferencia más evidente es el más reducido número de ponderaciones. En el modelo explicativo del análisis factorial, el investigador no puede controlar las ponderaciones. En el modelo de medida sin embargo, el investigador tiene un control completo sobre las variables descritas por cada constructo. En el ejemplo, cada variable era un indicador de un solo constructo; por tanto, existe un número más reducido de ponderaciones. Aunque una variable puede ser indicador de más de un constructo, este método no es recomendable excepto en situaciones específicas con fuerte carga teórica. El investigador especifica un modelo de medida tanto para los constructos exógenos como para los constructos endógenos exactamente de esta forma.

Determinación del número de indicadores Ya hemos discutido la lógica y la justificación del uso de indicadores múltiples (variables) para representar un constructo, pero queda una cuestión fundamental: ¿cuántos indicadores deberían utilizarse por constructo? El número mínimo de indicadores por constructo es uno, pero el uso de sólo un único indicador exige al investigador dar estimaciones de fiabilidad. Un constructo puede ser representado por dos indicadores, pero tres es el número mínimo preferido de indicadores, porque utilizando sólo dos indicadores aumentan las oportunidades de alcanzar una solución no factible [36]. Además de los fundamentos teóricos que deberían utilizarse para seleccionar las variables como indicadores de un constructo, no existe límite superior en términos de número de indicadores. A efectos prácticos, sin embargo, la mayoría de los constructos deberían ser representados por un número de entre cinco y siete indicadores. La excepción notable es el uso de escalas preexistentes, que pueden contener muchos ítems, cada uno actuando como indicador del constructo. En estos casos, el investigador debería evaluar la **unidimensionalidad** del constructo y la posibilidad de múltiples subdimensiones que puedan ser representadas en un modelo factorial de segundo-orden (véase sexto paso para una discusión más extensa).

Justificar la fiabilidad del constructo Una vez que el modelo de medida ha sido especificado, el investigador debe probar después la fiabilidad de los indicadores. Ofreceremos una discusión de la fiabilidad mucho más detallada en el paso 6, pero en este paso el investigador debería determinar el método básico en que se establece la fiabilidad de cada constructo. Existen dos métodos principales para establecer la fiabilidad: (1) estimación empírica o (2) especificación por el investigador.

Fiabilidades estimadas empíricamente La estimación empírica de la fiabilidad sólo es posible si el constructo tiene dos o más indicadores. Para un constructo con un sólo indicador, el investigador debe especificar la fiabilidad. Para la estimación empírica, el investigador especifica la matriz de ponderaciones tal y como se ha descrito, junto con un término de error para cada variable indicador (dado que no esperamos predecir cada indicador perfectamente). Cuando los modelos de medida y estructural están estimados, los coeficientes de ponderación ofrecen estimaciones de las fiabilidades de los indicadores y del constructo conjunto. En este enfoque, la intervención del investigador no tiene impacto sobre el valor de la fiabilidad utilizado en la estimación del modelo excepto en los conjuntos de indicadores incluidos. Ilustramos los pasos exactos exigidos para este enfoque en el sexto paso.

Especificar las fiabilidades En algunos casos resulta apropiado especificar o «fijar», las fiabilidades. La especificación de las fiabilidades para el indicador(es) de cualquier constructo latente puede parecer que va en contra de los objetivos de la modelización de ecuaciones estructurales; sin embargo, en al menos tres situaciones está justificado y es altamente recomendable. En un caso,

la estimación empírica de la fiabilidad no es posible, aunque el investigador puede saber que el error de medida todavía existe. En otros, los indicadores pueden haber sido utilizados previamente; por tanto, se conocen las fiabilidades antes de utilizarse. Y finalmente, tenemos un enfoque en dos pasos en el que las fiabilidades se evalúan en primer lugar y a continuación se especifican en el proceso de estimación. Este enfoque de dos pasos separa explícitamente los dos procesos empíricos y ofrece perspectivas de los dos por separado.

Medidas de ítem único Con medidas de ítem único, no es posible estimar empíricamente la fiabilidad; por tanto, el investigador se enfrenta con dos posibilidades. La primera, establecer («fijar») la fiabilidad a 1,0, indicando que no existe error de medida en el indicador. Pero como ya hemos discutido, sabemos que esto es erróneo en casi todos los casos, si por ninguna otra razón la fiabilidad está afectada por la calidad de la recogida de datos. Por ejemplo, el género puede ser «perfecto» o muy cercano (99 por ciento), con errores debidos a errores de codificación. Sin embargo, los ingresos pueden tener un nivel elevado de error (es decir, un 10 por ciento) debido a sesgos de las contestaciones y al nivel de medida. Muy a menudo por tanto, el investigador debería realizar alguna estimación de la fiabilidad y especificar el valor de los indicadores de ítem único. En Hayduk [54] se ofrecen varios enfoques recomendados.

Uso de medidas o escalas validadas Muchas veces el investigador emplea una escala o medida que ha sido extensamente contrastada en la investigación previa. Si al usarla el objetivo es obtener una réplica de los efectos encontrado en estudios anteriores, entonces la fiabilidad de la medida o escala debería fijarse en niveles previamente especificados. Este es un ejemplo de las fiabilidades especificadas por el investigador para mantener el control sobre el significado de los constructos. Al fijar la fiabilidad, el investigador «fuerza» a un indicador para tener la cantidad de varianza apropiada para el constructo y mantener un significado específico para el constructo.

Análisis de dos pasos Muchos investigadores proponen un proceso de dos pasos de modelización de ecuaciones estructurales en el que el modelo de medida se «fija» en el segundo paso cuando el modelo estructural está estimado [4, 61, 68, 77, 102]. La razón de este enfoque es que la representación precisa de la fiabilidad de los indicadores se realiza mejor en dos pasos para evitar la interacción de los modelos de medida y estructurales. Aunque no podemos evaluar verdaderamente los modelos de medida y estructural aisladamente, debemos considerar el potencial para dentro del constructo frente a los efectos entre constructos en estimación, que pueden ser sustanciales y resultar en lo que Burt [28] denomina «confusión interpretacional».

Un análisis de dos pasos con estimación simultánea tanto de modelos de medida como estructurales es la mejor aproximación cuando el modelo posee tanto fuerte lógica teórica como medidas altamente fiables, resultando en relaciones más precisas y disminuyendo la posibilidad de interacción. Sin embargo, cuando nos enfrentamos con medidas que son menos fiables, o una teoría que es sólo tentativa, el investigador debería considerar un enfoque por pasos para maximizar la interpretabilidad tanto de las medidas como de los modelos estructurales. Se ha producido un debate considerable sobre la conveniencia de este enfoque y sobre aquellos casos en los que está justificado, tanto conceptual como empíricamente [5, 44, 45, 55].

Métodos para especificar la fiabilidad Para «fijar» la fiabilidad de un indicador en una matriz de correlación, el investigador especifica el valor de ponderación como la raíz cuadrada de la fiabilidad deseada o estimada, o especifica el término de error de esa variable como 1,0 menos el valor de fiabilidad deseado. Si se utiliza la matriz de covarianza, entonces el término de error o valor de ponderación se multiplica por el valor de la varianza de la variable. Al especificar las fiabilidades, el investigador puede determinar los valores de carga, el término de error o ambos. Dado que especificar bien la ponderación o bien los términos de error determina automáticamente el otro valor, recomendamos que ambos se fijen para el modelo de mayor parsimonia y que un coeficiente no se utilice para estimar un valor que podría ser especificado. Este procedimiento pue-

de ser realizado simplemente en programas estadísticos a través de un tratamiento único o para cada variable.

Una variante de la fijación de la fiabilidad de la escala completa es un enfoque propuesto por Hayduk [54, 55] que implica un proceso por pasos. En primer lugar, el investigador selecciona el indicador aislado que se cree es la mejor representación del constructo. Para este indicador, la fiabilidad (ponderación y término de error) es fija. A continuación, pueden añadirse indicadores adicionales y sus ponderaciones y términos de error estimados, permitiendo el cálculo de la fiabilidad del constructo completo. La lógica que hay detrás de este enfoque es especificar el significado deseado del constructo a través de este indicador y a continuación permitir a otros indicadores añadir significados al concepto ya definido. El inconveniente es que cuando todas las ponderaciones y los términos de error se estiman empíricamente de forma simultánea, el investigador no está seguro exactamente de qué constructo representa excepto algún concepto subyacente común al resto de los indicadores.

Correlaciones entre constructos e indicadores

Además de los modelos de medida y estructurales, el investigador también especifica cualquier correlación entre los constructos exógenos o entre los constructos endógenos. Muchas veces los constructos exógenos están correlacionados, representando una influencia «compartida» sobre las variables endógenas. Las correlaciones entre los constructos endógenos, sin embargo, tienen pocas aplicaciones apropiadas y no están recomendadas para un uso normal debido a que representan correlaciones entre las ecuaciones estructurales que pueden confundir sus interpretaciones. Finalmente, los indicadores en el modelo de medida también pueden estar correlacionadas separadamente de las correlaciones de los constructos. Este método se suele evitar excepto en situaciones específicas, tales como un estudio en que existen efectos conocidos del proceso de recogida de datos o de medida sobre dos o más indicadores, o un estudio longitudinal en el que el mismo indicador se recoge en dos períodos de tiempo [8, 89]. Para una discusión más completa de los efectos distorsionadores de los indicadores correlacionados, véase Gerbing y Anderson [47].

Cuarto paso: Selección del tipo de matriz de entrada y estimación del modelo propuesto

Como se muestra en los pasos precedentes, se exige mucho más al investigador que utiliza SEM en términos de especificación del modelo a estimar que con otra técnica multivariante, con la posible excepción del análisis conjunto. Ahora el investigador debe llevar el proceso efectivo de estimación del modelo especificado, incluyendo los supuestos de entrada de datos en la forma apropiada y selección del procedimiento de estimación (véase Figura 11.6). Las decisiones tomadas en estas áreas tienen un impacto profundo en los resultados conseguidos.

Introducción de datos

SEM difiere de otras técnicas multivariantes en que utiliza sólo la matriz de varianza-covarianza o de correlación como sus datos de entrada. Las observaciones individuales pueden ser introducidas en los programas, pero se convierten en uno de estos dos tipos de matrices antes de la estimación. El interés en SEM no está en las observaciones individuales sino el patrón de relaciones entre los encuestados. La entrada del programa es una matriz de correlación o varianza-covarianza de todos los indicadores utilizados en el modelo. El modelo de medida especifica entonces qué indicadores corresponden a cada constructo, y las puntuaciones del constructo latente son las empleadas en el modelo estructural.

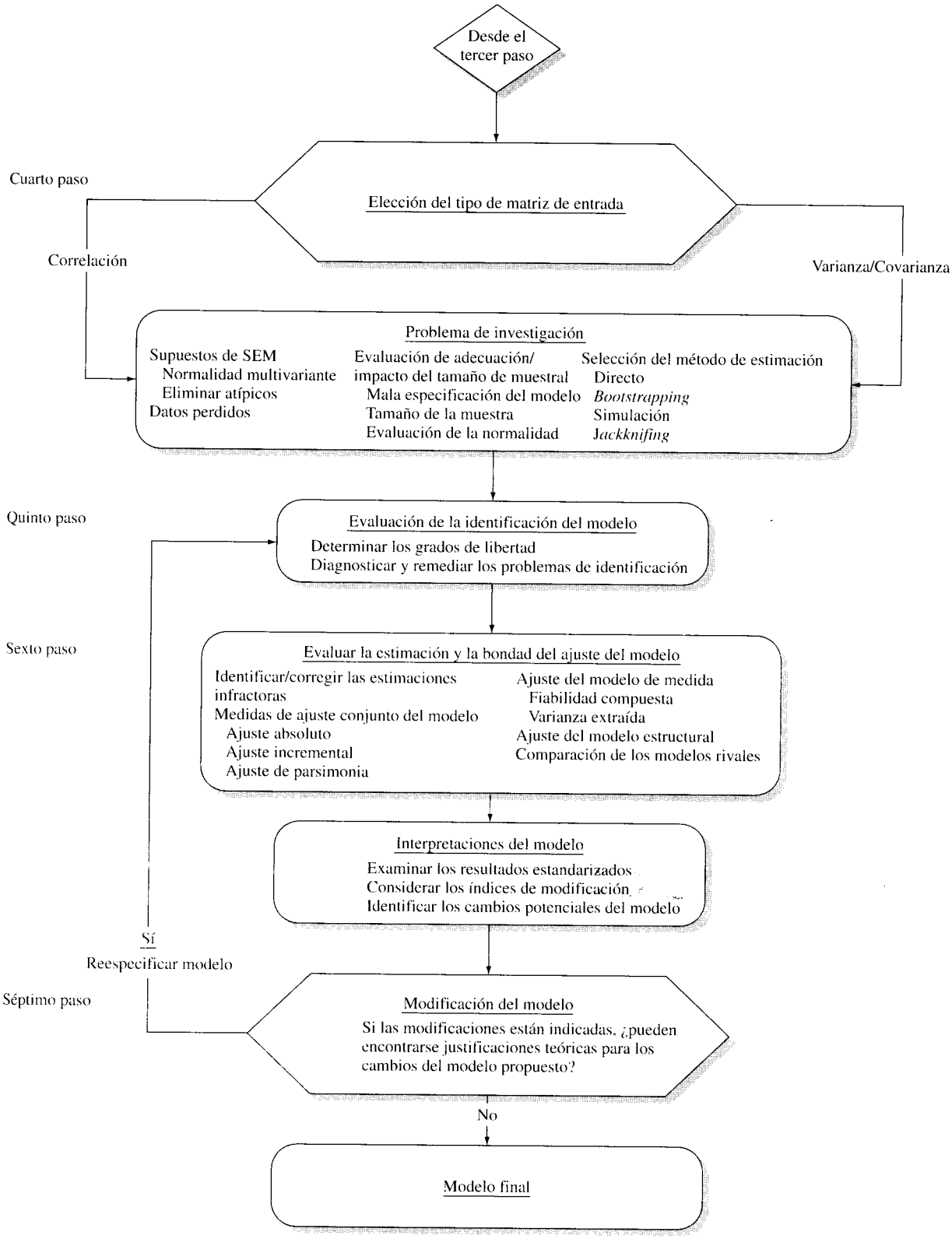


FIGURA 11.6. Pasos 4-7 en un proceso de siete pasos para la modelización de ecuaciones estructurales.

Supuestos SEM comparte tres supuestos con los otros métodos multivariantes que hemos estudiado: observaciones independientes, muestra aleatoria de los encuestados y la linealidad de todas las relaciones. Además, SEM es más sensible a las características distribucionales de los datos, particularmente el incumplimiento de la normalidad multivariante (crítica en el uso de LISREL) o una fuerte curtosis (apuntamiento) de los datos. Algunos programas informáticos, como EQS, son menos sensibles a los datos no normales, pero los datos deberían ser evaluados sin importar qué programa se está utilizando. Los mínimos cuadrados generalizados, un método de estimación alternativo, pueden ajustarse a estas violaciones, pero este método se hace rápidamente impracticable a medida que el modelo aumenta en tamaño y complejidad; por tanto su uso es limitado [98]. La ausencia de normalidad multivariante es particularmente problemática porque infla sustancialmente el estadístico de la chi-cuadrado y crea un sesgo alcista en valores críticos para determinar la significación de los coeficientes [78, 98].

Dado que los programas sólo aceptan las matrices de varianza-covarianza o correlación, el investigador debe realizar todos los test de diagnóstico sobre los datos antes de utilizarlos en el procedimiento de estimación. Aunque los programas de ecuaciones estructurales no tienen procedimientos de diagnóstico para contrastar estos tres supuestos, pueden ser contrastados con métodos convencionales como PRELIS [66]. El investigador también debería identificar cualquier atípico en los datos antes de convertirlos a forma matricial.

Datos ausentes Los datos ausentes pueden tener un profundo efecto sobre el cálculo de la matriz de entrada de datos y su capacidad para ser utilizados en el proceso de estimación. Existen dos formas en las que se pueden incorporar los valores perdidos en SEM. La primera es el método directo, en el que los parámetros del modelo se estiman tanto con datos completos como incompletos [2, 67]. Esta aproximación se utiliza raramente, sin embargo, dada la complejidad del modelo resultante. Más habitual es el método indirecto, para el cual se estima una matriz de datos de entrada utilizando toda o parte de la información disponible. Como se discutía en el Capítulo 2, existen muchos métodos disponibles para «resolver» el problema de los datos ausentes, que van desde la supresión según el orden de lista a los métodos de imputación. La investigación reciente ha mostrado que el método EM introduce menos sesgo en los modelos estimados, pero que las opciones pareada y en forma de lista se desarrollan adecuadamente si la proporción de datos ausentes no es muy elevada [26]. Una conclusión del método de listas es que puede reducir seriamente el tamaño muestral, un tema clave como veremos en la discusión posterior. La aproximación pareada puede introducir irregularidades en la matriz de datos de entrada que provocarán después serios problemas en el proceso de estimación. Por tanto, nos existe un método único que produzca siempre los mejores resultados, y si es posible que el investigador debiera emplear varias aproximaciones para evaluar la estabilidad de los resultados.

Covarianzas versus correlaciones Un tema importante en la interpretación de los resultados es el uso de la matriz de varianza-covarianza frente a la matriz de correlación. SEM se formuló inicialmente para ser utilizada con la matriz de varianza-covarianza (de ahí su denominación habitual como análisis estructural de la covarianza). La matriz de covarianza tiene la ventaja de proporcionar comparaciones válidas entre diferentes poblaciones o muestras, característica que no es posible cuando los modelos se estiman con una matriz de correlación. La interpretación de los resultados, sin embargo, es de alguna forma más difícil cuando se utilizan covarianzas dado que los coeficientes deben interpretarse en términos de las unidades de medida de los constructos.

La matriz de correlación tiene un uso muy extendido en muchas aplicaciones. Las matrices de correlación tiene un rango común que hace posible las comparaciones directas de los coeficientes dentro de un modelo, dado que simplemente es una matriz de varianza-covarianza «estandarizada» en la que la escala de medida de cada variable se modifica dividiendo las varianzas o covarianzas por el producto de las desviaciones estándar. Utilizar las correlaciones es apropiado cuando el objetivo del investigador es sólo entender el patrón de las relaciones entre constructos, pero no

explicar la varianza total del constructo. Otro uso apropiado es hacer comparaciones entre las diferentes variables, dado que las escalas de medida afectan a las covarianzas. Los coeficientes obtenidos de la matriz de correlación están siempre en unidades estandarizadas, similares a las ponderaciones beta de la regresión, y con un rango entre −1,0 y +1,0. Además, se ha demostrado que la matriz de correlación ofrece estimaciones más conservadoras de la significación de los coeficientes y no están sesgadas al alza, como se pensaba anteriormente [35].

En resumen, el investigador debería emplear una matriz de varianza-covarianza siempre que se desarrolle un «contraste de la teoría» real, en el que las varianzas y las covarianzas satisfacen los supuestos y la metodología y son la forma apropiada de los datos para validar las relaciones causales. Sin embargo, a menudo el investigador se centra sólo en las pautas de las relaciones, no con la explicación total tal y como se necesita en el contraste de la teoría, siendo aceptable la matriz de correlación. Si se utiliza la matriz de correlación, el investigador debería interpretar con cautela los resultados y su generalización a diferentes situaciones.

Tipos de correlaciones o covarianzas utilizadas Uno de los modos más extendidos de cálculo de las correlaciones o covarianzas entre las variables observadas es la correlación de producto-momento de Pearson. También es la forma más habitual de correlación utilizada en el análisis multivariante, simplificando al investigador el cálculo de las matrices de varianzas y covarianzas. El supuesto de la correlación del momento-producto es que ambas variables se miden métricamente. Esto hace la correlación momento-producto inadecuada para el uso con medidas no métricas (ordinales o binarias). Para permitir la incorporación de las medidas no métricas en modelos de ecuaciones estructurales, el investigador debe emplear diferentes tipos de correlación. Si ambas variables son ordinales con tres o más categorías (policotómicas), entonces es apropiada la **correlación policórica**. Si las medidas no métricas son binarias, entonces se utiliza la **correlación tetracórica**. Para casos en los que la medida métrica está relacionada con una medida ordinal policotómica, la **correlación poliserial** es la que representa la relación. Si una medida binaria está relacionada con una medida métrica, se utiliza una **correlación biserial**.

Tamaño muestral Incluso aunque las observaciones individuales no sean necesarias, como sí ocurría con el resto de los métodos multivariantes, el tamaño muestral juega un papel importante en la estimación e interpretación de los resultados SEM. El tamaño muestral, como con cualquier otro método estadístico, ofrece bases para la estimación del error muestral. La cuestión crítica en SEM es el tamaño muestral necesario. Aunque no existe ningún criterio que dicte el tamaño de muestra necesario, existen al menos cuatro factores que afectan a los requisitos del tamaño muestral [83]: (1) mala especificación del modelo, (2) tamaño del modelo, (3) no cumplimiento de la normalidad, y (4) procedimiento de estimación.

Mala especificación del modelo La mala especificación del modelo se refiere a la medida en que el modelo sufre por los errores de especificación. Como ya se ha discutido, el error de especificación es la omisión de variables relevantes del modelo especificado. Todos los modelos de ecuaciones estructurales sufren de errores de especificación en la medida en que no puede ser incluido cada indicador o constructo potencial. El impacto de los constructos e indicadores omitidos, sin embargo, debería de ser insignificante si el investigador ha incluido todos aquellos relevantes según la teoría. El tamaño muestral afecta a la capacidad del modelo para ser estimado correctamente e identificar el error de especificación si se desea. Por tanto, si el investigador está preocupado por el impacto del error de especificación, las exigencias sobre el tamaño muestral deberían aumentarse por encima de aquellos que se hubiesen requerido en otro modelo.

Tamaño del modelo El tamaño de muestra mínimo absoluto debe ser al menos tan grande como el número de covarianzas y correlaciones de la matriz de datos de entrada. Sin embargo, más habitual es un mínimo de al menos cinco encuestados para cada parámetro estimado, con un ratio de 10 encuestados por parámetro considerado más apropiado. Por tanto, a medida que la complejidad

del modelo aumenta, también aumentan los requisitos. Nótese que estos requisitos difieren del concepto de grados de libertad (discutido más tarde) y hacen referencia al número de encuestados originales utilizado para calcular la matriz de correlación o covarianza.

Incumplimientos de la normalidad En la medida en que los datos no cumplan los supuestos de normalidad multivariante, el ratio de los encuestados respecto de los parámetros necesita aumentar a una tasa generalmente aceptada de 15 encuestados para cada parámetro. Aunque algunos procedimientos de estimación se diseñan específicamente para tratar con datos no normales, siempre se anima al investigador a ofrecer un tamaño de muestra suficiente para que el impacto de error muestral sea mínimo, especialmente para datos no normales [98].

Procedimiento de estimación Se ha demostrado que el método de **Estimación máximo verosimil (MLE)**, el procedimiento de estimación más habitual, ofrece resultados válidos con muestras de tan sólo 50 observaciones, pero una muestra de este tamaño *no es recomendable*. Se acepta generalmente que el tamaño de muestra mínimo que asegura el uso apropiado de MLE es de 100 a 150 [36]. A medida que aumentamos el tamaño muestral por encima de ese valor, el método MLE aumenta en su sensibilidad al tamaño muestral por encima de ese valor. A medida que el tamaño muestral aumenta (pasando de 400 a 500), el método se hace «más sensible» y casi no se detecta ninguna diferencia, haciendo que todas las medidas de calidad del ajuste indiquen un ajuste muy pobre [30, 73, 94]. Aunque no existe un tamaño muestral correcto, se recomiendan tamaños que vayan entre 100 y 200. Un enfoque es contrastar siempre un modelo con un tamaño de muestra de 200, sin importar cuál fuera el tamaño muestral original, debido a que algunos investigadores han propuesto 200 como un «tamaño de muestra crítico» [58]. Si se selecciona un procedimiento de estimación asintóticamente libre de distribución (véase la siguiente sección para más detalles), los requisitos del tamaño muestral aumenta sustancialmente, ya que exige mayores tamaños de muestra para compensar la dependencia sobre los supuestos de distribución mediante otros métodos.

Resumen Existen muchos factores que afectan al tamaño muestral exigido. Normalmente, recomendamos un tamaño muestral de 200, aumentando si se sospecha que existe mala especificación, el modelo es muy grande o complejo, los datos exhiben características no normales o se utiliza un procedimiento de estimación alternativo. Una evaluación aislada es el diagnóstico de la *N* crítica, que es el tamaño de muestra que haría significativo el nivel de ajuste del modelo (medido por χ^2), llegando al nivel especificado de significación [58]. La medida chi-cuadrado está disponible en la mayoría de los programas y es una base para la comparación del tamaño muestral. Se anima al investigador, si duda, a determinar la sensibilidad del modelo a variaciones en el tamaño muestral estimando el modelo bajo especificaciones distintas del tamaño muestral.

Estimación del modelo

Una vez que están especificados los modelos estructurales y de medida y que se ha seleccionado el tipo de datos de entrada, el investigador debería elegir cómo se estimará el modelo. En el caso de SEM, tiene varias opciones tanto para el procedimiento de estimación como respecto al programa informático a utilizar.

Técnica de estimación Los intentos iniciales de estimación del modelo de ecuaciones estructurales se realizaron con la regresión de los mínimos cuadrados ordinarios (OLS). Pero estos esfuerzos fueron superados rápidamente por la estimación máximo verosímil, que es eficiente y no sesgada cuando se cumplen los supuestos de normalidad multivariante. Como tal, fue

utilizada en versiones iniciales de LISREL y se ha convertirdo en una técnica ampliamente empleada en la mayoría de los programas informáticos. La sensibilidad de MLE a la no normalidad, sin embargo, creó una necesidad de técnicas de estimación alternativas y pronto aparecieron métodos tales como los mínimos cuadrados ponderados (WLS), mínimos cuadrados generalizados (GLS) y asintóticamente libre de distribución (AGL) [55]. La técnica AGL ha recibido recientemente atención particular debido a su insensibilidad a la no normalidad de los datos. Su principal conclusión es la exigencia de un aumento del tamaño de muestra. Todas las técnicas de estimación alternativas han empezado a difundirse ampliamente a medida que ha aumentado el uso de ordenadores personales, haciéndoles factibles resolver cualquier problema habitual.

Procesos de estimación Además de la técnica de estimación empleada, el investigador puede también escoger entre varios procesos de estimación. Estos procesos van desde la estimación directa del modelo, que es similar a lo que hemos visto en el resto de las técnicas multivariantes, a métodos que generan miles de estimaciones del modelo para las cuales se obtienen los resultados finales del modelo. En este capítulo discutiremos cuatro procesos —estimación directa, *bootstrapping*, simulación y análisis *jackknife*—. Se remite al lector al Capítulo 12 para una discusión más detallada de varios de estos procesos.

Estimación directa El proceso de estimación más habitual es el de **estimación directa**, en el cual se estima directamente un modelo con un procedimiento de estimación elegido. En este proceso, como el modelo multivariante de los capítulos anteriores, estimamos el parámetro; a continuación el intervalo de confianza (y error estándar) de cada parámetro estimado se basa en el error muestral. Tanto los parámetros estimados como su intervalo de confianza provienen del modelo estimado de una muestra aislada.

Bootstrapping Sin embargo, existen varias alternativas que no descansan en un único modelo de estimación sino en estimaciones del parámetro calculado y sus intervalos de confianza basados en estimaciones múltiples [24]. La primera opción es **bootstrapping**, que se realiza en cuatro pasos. En el primero, se diseña la muestra original para que actúe como la población a efectos muestrales. En el segundo paso, se vuelve a muestrear la muestra original un número especificado de veces (quizá hasta varios miles) para generar un gran número de nuevas muestras, siendo cada una un subconjunto aleatorio de la muestra original. En el tercer paso, se estima el modelo y para cada muestra nueva se guardan los parámetros estimados. En el último paso, las estimaciones de los parámetros finales se calculan como la media de las estimaciones de los parámetros de todas las muestras. El intervalo de confianza no se estima por un error muestral, sino que se observa directamente examinando la distribución efectiva de los parámetros estimados alrededor de la media. De esta forma, las estimaciones del parámetro final y sus intervalos de confianza se obtienen directamente de las estimaciones de modelos múltiples para varias muestras aisladas y no descansan en supuestos como la distribución estadística de los parámetros.

Simulación El investigador puede emplear técnicas de **simulación**, que también descansan en muestras múltiples y modelos estimados. Los procesos de simulación difieren de *bootstrapping* en que durante el proceso de generación de nuevas muestras, el programa de simulación puede cambiar ciertas características de la muestra para cumplir con lo que se propone el investigador. Por ejemplo, el grado de correlación entre variables puede ser variable a lo largo de las muestras en cierta forma sistemática. De esta forma, el investigador no sólo tiene una variación asintótica de la muestra entre las muestras sino también una pauta sistemática especificada en el procedimiento de simulación. Con estas muestras, se estiman otra vez los modelos para cada muestra y los resultados compilados, como en el proceso de *bootstrapping*.

Análisis de *Jackknife* El último proceso de estimación es el procedimiento de **jackknife**, donde de nuevo se crean muestras repetidas a partir de la muestra original. El *jackknife* difiere de los procedimientos de simulación y *bootstrapping*, sin embargo, en el método de creación de nuevas muestras. En lugar de crear un gran número de nuevas muestras aleatorias, el proceso de *jackknife* crea N nuevas muestras, donde N es el tamaño muestral original. Cada vez que se crea una nueva muestra, se omite una observación diferente. Por tanto, cada nueva muestra tiene un tamaño de muestra de $N - 1$ con una observación diferente omitida en cada muestra. La ventaja de este proceso es la facilidad de identificar **observaciones influyentes** mediante el examen de los cambios en los parámetros estimados. Si se desea, la estimación del último parámetro puede calcularse como el parámetro medio, pero en los casos de tamaños de muestras reducidas no existen nuevas muestras para calcular adecuadamente el intervalo de confianza.

Programas informáticos Una vez que se ha elegido el procedimiento de estimación, el siguiente paso es escoger el programa informático utilizado para estimar de forma efectiva el modelo. El programa más utilizado es LISREL (*Linear Structural RELations*) [64, 65, 67], un modelo verdaderamente flexible para varias situaciones de investigación (sección-cruzada, experimental, cuasi-experimental y estudios longitudinales). LISREL ha encontrado aplicaciones a lo largo de todos los campos de estudio [8] y se ha convertido casi en sinónimo de la modelización de ecuaciones estructurales. Sin embargo, existen varios programas informáticos alternativos, entre ellos EQS [13, 14, 18], AMOS [6], PROC CALIS de SAS [53], COSAN [46] o LVPLS [104]. EQS tiene supuestos menos exigentes sobre la normalidad multivariante de los datos y LVPLS se ajusta mejor para la predicción aunque es limitado a efectos de interpretación de los resultados. AMOS se ha ganado una creciente popularidad en los últimos años debido a la sencillez del interfaz para el usuario y se ha comparado recientemente con LISREL y EQS [59]. Todos estos programas se encuentran disponibles en versiones que pueden ser utilizadas por ordenadores personales.

Resolución del problema de «definida no positiva»

Un problema habitual para todos los usuarios de SEM es el mensaje de error del ordenador, «la matriz... es definida no positiva». Lo que ha ocurrido es que tanto la matriz de datos de entrada como la matriz de datos estimados es singular, lo que significa que existe una dependencia lineal o inconsistencia entre algún grupo de variables. Existen muchas causas de este problema, pero se han encontrado algunas generales [105]. Si el error se produce en la matriz de entrada de datos, las causas más probables son (1) el enfoque para tratar con los datos ausentes utilizado, especialmente la eliminación pareada [6]; o (2) una dependencia lineal entre las variables, incluyendo todos los ítems de la escala y el total de la escala en la matriz de entrada. En estos dos casos, el investigador debería generar una nueva matriz de datos, empleando un proceso de datos ausentes alternativo o eliminando las variables infractoras. Si el problema se produce en la matriz de datos estimados, entonces el investigador debería corregir cualquier varianza de error negativa (conocida como casos Heywood y que se describen en el paso 6) o intentar valores de entrada distintos.

Quinto paso: Valoración de la identificación del modelo estructural

Durante el proceso de estimación, la causa más probable de la «parada» del programa de ordenador o la producir resultados sin sentido o ilógicos es un problema de identificación del modelo estructural. Un problema de **identificación**, en términos sencillos, es la incapacidad del modelo propuesto para generar estimaciones aisladas. Se basa en el principio de que debemos tener una ecuación aislada y diferenciada para estimar cada coeficiente, reflejado en el dicho de que «se deben tener más ecuaciones que incógnitas» que aprendimos en álgebra cuando definíamos un sis-

tema de ecuaciones. Sin embargo, a medida que el modelo estructural se hace más complejo, no existe un enfoque garantizado para asegurar que el modelo está identificado [23].

Grados de libertad

A efectos de identificación, el investigador se centra en el tamaño de las matrices de correlación o covarianzas relativas respecto al número de coeficientes estimados. La diferencia entre el número de correlaciones o covarianzas y el número efectivo de coeficientes en el modelo propuesto se denomina **grados de libertad**. Al igual que los grados de libertad que encontramos en la regresión múltiple o MANOVA, un grado de libertad es un elemento no restringido de la matriz de datos. El número de grados de libertad para un modelo propuesto se calcula como

$$gl = \frac{1}{2}[(p+q)(p+q+1)] - t$$

donde

p = número de indicadores endógenos
q = el número de indicadores exógenos
t = número de coeficientes estimados en el modelo propuesto

La primera parte de la ecuación calcula el tamaño no redundante de la matriz de covarianza o correlación (es decir, la mitad inferior o superior de la matriz más la diagonal). Entonces cada coeficiente estimado «gasta» un grado de libertad. La principal diferencia entre los grados de libertad utilizados en SEM comparado con otras técnicas multivariantes es que el número de parámetros estimados se compara con el número de elementos de la matriz de datos, no con el tamaño muestral. En SEM el tamaño muestral se utiliza para estimar el error de muestreo, pero *no afecta a los grados de libertad.*

Normas para la identificación

Aunque no existe una regla aislada que establecerá la identificación de un modelo, el investigador dispone de varias «normas» o heurística [10, 34, 84, 85]. Las dos normas más básicas son las condiciones de orden y rango. Las **condiciones de orden** afirman que los grados de libertad del modelo deben ser mayores o iguales a cero. Esto corresponde a lo que hemos denominado como modelo identificado o modelo sobreidentificado. Un **modelo identificado** tiene exactamente cero grados de libertad. Aunque esto ofrece un ajuste perfecto del modelo, la solución no tiene interés puesto que no se puede generalizar. Un **modelo sobreidentificado** es el objetivo de todos los modelos de ecuaciones estructurales. Tiene más información en la matriz de datos que el número de parámetros a estimar, lo que significa que tiene un número positivo de grados de libertad. Al igual que en otras técnicas multivariantes, el investigador se esfuerza por conseguir un ajuste aceptable con el mayor grado de libertad posible. Esto asegura que el modelo es tan generalizable como sea posible. Un modelo que no llega a cumplir la condición de orden se conoce como un **modelo infraestimado**. Este modelo tiene grados de libertad negativos, lo que significa que se intentan estimar más parámetros de lo que permite la información disponible. El modelo no puede ser estimado hasta que algunos parámetros sean fijos o restringidos.

La condición de orden es una condición de identificación necesaria, pero no suficiente. El modelo también debe cumplir la **condición de rango**, lo que exige que el investigador determine algebraicamente si cada parámetro se identifica (estima) especialmente. Pero incluso para los modelos más sencillos, se trata de un ejercicio muy complejo para ser considerado directamente por el investigador. En su lugar, hay heurística disponible. Primero está la norma de las tres medidas, que evalúa que cualquier constructo con tres o más indicadores siempre estará identificado. También existe la norma del modelo recursivo, que dice que los modelos recursivos con constructos

identificados (norma de las tres medidas) siempre estarán identificados [85]. Un modelo recursivo no tiene relaciones no recursivas o recíprocas en el modelo estructural.

Diagnóstico de los problemas de identificación

Los programas de ecuaciones estructurales también realizan contrastes para diagnosticar problemas de identificación. LISREL tiene un test sencillo de identificación durante el proceso de estimación para examinar la matriz de información, mientras que EQS tiene el test de rango de Wald. Aunque estos contrastes identifican la mayoría de los problemas de identificación, pueden no evaluar la unicidad de cada parámetro estimado, como exige la condición de rango.

El investigador puede llevar a cabo contrastes cuando la ecuación se identifica para ver si los resultados son inestables debido al nivel de identificación. En primer lugar, el modelo puede ser reestimado varias veces, cada una de ellas con un **valor de partida** diferente. El investigador puede especificar un valor inicial para cualquier parámetro estimado, un «punto de partida» del proceso de estimación. Si el valor de partida no está disponible, el programa lo calcula automáticamente por diversos métodos. Si los resultados no convergen en el mismo punto para diversos valores de partida, la identificación debería examinarse más a fondo. El segundo test evalúa el efecto de la identificación sobre un único coeficiente, que consiste en estimar en primer lugar el modelo y a continuación obtener la estimación de coeficiente. Posteriormente, se «fija» el coeficiente a su valor estimado y se reestima la ecuación (la fijación de los coeficientes se discute en una sección posterior). Si el ajuste conjunto del modelo varía significativamente, esto indica problemas de identificación.

Otro enfoque es observar los posibles síntomas de un problema de identificación. Estos incluyen (1) errores estándar muy elevados para uno o más coeficientes, (2) la incapacidad del programa para invertir la matriz de información, (3) estimaciones muy poco razonables o estimaciones imposibles tales como varianzas de error negativas, o (4) elevadas correlaciones ($\pm$ 0,90 o superiores) entre los coeficientes estimados. En el paso 6 se tratará de interpretar estos resultados con mayor detalle.

Causas y soluciones para los problemas de identificación

Si se localiza un problema de identificación, el investigador debería buscar tres posibles causas: (1) un mayor número de coeficientes estimados relativos al número de correlaciones o covarianzas, indicada por el reducido número de grados de libertad –similar al problema de sobreajuste de los datos encontrado en otras técnicas multivariantes; (2) el uso de efectos recíprocos (flechas causales de dos sentidos entre dos constructos); o (3) fallo en la fijación de la escala de un constructo. Discutiremos este procedimiento más adelante en nuestro análisis de los datos del ejemplo.

La única solución para un problema de identificación es definir más restricciones para el modelo –esto es, eliminar algunos de los coeficientes estimados. El investigador debería seguir un proceso estructurado, añadiendo gradualmente más restricciones (eliminando secuencias del diagrama de secuencias) hasta que el problema esté solucionado. Al hacerlo, el investigador está intentando conseguir un modelo sobreidentificado que tiene grados de libertad suficientes con las cuales evaluar, si es posible, la cantidad de error de medida y de muestreo y ofrecer mejores estimaciones de las «verdaderas» relaciones casuales. Para llegar a este fin, Hayduk recomendó el siguiente proceso [54]: (1) construir un modelo teórico con el mínimo número de coeficientes (incógnitas) que puedan ser justificados. Si se encuentran problemas de identificación, procederemos con las soluciones por este orden: (2) fijar las varianzas de error de medida de los constructos si es posible, (3) fijar algunos coeficientes estructurales que sean conocidos con fiabilidad, y (4) eliminar las variables problemáticas. Si los problemas de identificación todavía existen, el investigador debe reformular el modelo teórico para ofrecer más constructos relativos al número de relaciones causales examinadas.

Sexto paso: Evaluación de los criterios de calidad de ajuste

El primer paso de la evaluación de los resultados es una inspección inicial de las «estimaciones infractoras». Una vez que el modelo está establecido como para ofrecer estimaciones aceptables, a continuación debe evaluarse la calidad del ajuste a diversos niveles: en primer lugar para el modelo conjunto y a continuación para los modelos estructurales y de medida por separado.

Estimaciones infractoras

En primer lugar se examinan los resultados buscando **estimaciones infractoras**. Se trata de coeficientes estimados tanto en los modelos de medida como los estructurales que exceden los límites aceptables. Los ejemplo más normales de estimaciones infractoras son (1) varianzas de error negativas o varianzas de error no significativas para cualquier constructo, (2) coeficientes estandarizados que sobrepasan o están muy cerca de 1,0, o (3) errores estándar muy elevados asociados con cualquier coeficiente estimado. Si se encuentran estimaciones infractoras, el investigador debería resolver en primer lugar cada caso antes de evaluar cualquier resultado específico del modelo, en la medida en que cambios en una parte del modelo puedan tener efectos significativos sobre otros resultados.

Se han utilizado varios enfoques para la resolución de estos problemas en la discusión de los problemas de identificación. Si se corrigen los problemas de identificación y aún así, los problemas todavía se mantienen, existen otras soluciones. En el caso de varianzas de error negativas (también conocidas como **casos de Heywood**), una posibilidad es fijar las varianzas de los errores infractores a un valor positivo muy pequeño (0,005) [17, 35]. Aunque este remedio cumple los requisitos prácticos del proceso de estimación, sólo enmascara el problema subyacente y debe ser considerado cuando se interpreten los resultados. Si las correlaciones de la solución estandarizada exceden de 1,0, o dos estimaciones están altamente correlacionadas, entonces el investigador debería considerar la eliminación de uno de los constructos o debería asegurarse que se ha establecido entre los constructos una verdadera validez discriminante. En muchos casos, tales situaciones son el resultado de modelos ateóricos, establecidos sin la suficiente justificación teórica o modificados solamente a partir de consideraciones empíricas.

Ajuste global del modelo

Una vez que el investigador ha establecido que no existen estimaciones infractoras, el siguiente paso es evaluar el ajuste global del modelo con una o más medidas de calidad del ajuste. La **calidad del ajuste** mide la correspondencia entre la matriz de entrada real u observada (covarianza o correlación) con la que se predice mediante el modelo propuesto.

Al desarrollar cualquier modelo estadístico, el investigador debe ser precavido con el «sobreajuste» del modelo a los datos. Al discutir la regresión, mostramos que deberían mantenerse ciertas relaciones (quizá cinco a uno) entre el número de coeficientes estimados y el número de encuestados. Esta relación debería mantenerse también en SEM. El investigador debería esforzarse en tener un gran número de grados de libertad, si el resto no cambia. Al hacerlo así, el modelo consigue **parsimonia** –el objetivo de un mejor o mayor ajuste del modelo para cada coeficiente estimado. Cuanto mejor ajuste pueda conseguirse con pocos coeficientes, mejor será el contraste del modelo y más confianza podemos tener en que los resultados no sean producto del sobreajuste de los datos.

Las medidas de calidad del ajuste son de tres tipos: (1) **medidas absolutas del ajuste**, (2) **medidas del ajuste incremental**, o (3) **medidas de ajuste de parsimonia**. Las **medidas absolutas del ajuste** evalúan sólo el ajuste global del modelo (tanto los modelos de medida como los estructurales colectivamente), sin ajuste para el grado de «sobreajuste» que pudiera ocurrir. Las **medidas del ajuste incremental** comparan el modelo propuesto con otro modelo especificado por el investigador. Finalmente, las **medidas de ajuste de parsimonia** «ajustan» las medidas de

ajuste para ofrecer una comparación entre modelos con diferentes números de coeficientes estimados, siendo el propósito determinar la cantidad del ajuste conseguido por cada coeficiente estimado.

El investigador se enfrenta con la cuestión de qué medidas elegir. No surge ninguna medida aislada o conjunto de medidas como las únicas medidas necesarias. Como SEM ha evolucionado en los últimos años, las medidas de calidad del ajuste se han estado desarrollando continuamente y se han propuesto medidas adicionales (no discutidas aquí) [78]. Se anima al investigador a emplear una o más medidas de cada tipo. La evaluación de la calidad del ajuste de un modelo es más un proceso relativo que un criterio absoluto. La aplicación de diversas medidas del ajuste múltiple permitirá al investigador ganar en consenso tanto para los tipos de medidas como para la aceptación del modelo propuesto. El Apéndice 11B describe con más detalle varias medidas dentro de cada clase de medidas de calidad del ajuste, incluyendo cómo calcular medidas no ofrecidas por el programa informático. Se puede encontrar una excelente revisión de las diversas medidas de la calidad del ajuste y su aplicación en múltiples situaciones en varias fuentes [22, 67]. Un nivel aceptable de calidad del ajuste conjunto no garantiza que todos los constructos cumplan los requisitos del ajuste del modelo, ni que el modelo estructural puede ser mantenido con certeza absoluta. El investigador debería evaluar cada una de estas áreas separadamente para confirmar su cumplimiento de los requisitos o como medio de identificar problemas potenciales que afectan a la calidad del ajuste conjunto del modelo.

Ajuste del modelo de medida

Una vez que se ha evaluado el ajuste del modelo conjunto, podemos evaluar la unidimensionalidad y la fiabilidad de la medida de cada constructo. La **unidimensionalidad** es un supuesto que subyace al cálculo de la fiabilidad y se demuestra cuando los indicadores de un constructo tienen un ajuste aceptable sobre un modelo de un único factor (uni-dimensional). Este tema está más allá del alcance de esta discusión, pero el lector interesado puede consultar diversas fuentes [3, 4, 5, 44, 45]. El uso de medidas de fiabilidad, tales como la **alfa de Cronbach** [32], no asegura la unidimensionalidad sino que en su lugar supone que existe. Se anima al investigador a que realice contrastes de unidimensionalidad sobre todos los constructos de indicadores múltiples antes de evaluar su fiabilidad. El siguiente paso es examinar las ponderaciones estimadas y evaluar la significación estadística de cada una. Si la significación estadística no se consigue, el investigador puede desear eliminar el indicador o intentar transformarlo para un mejor ajuste del constructo.

Fiabilidad compuesta Antes de realizar el examen de las ponderaciones de cada indicador, se utiliza una medida fundamental para la evaluación del modelo de medida, la fiabilidad compuesta de cada constructo. La fiabilidad es una medida de la consistencia interna de los indicadores del constructo, que representa el grado en que éstos «indican» el constructo común latente (no observado). Medidas más fiables ofrecen al investigador una mayor confianza de que todos los indicadores individuales son consistentes en sus medidas. Un valor umbral comúnmente aceptado para aceptar la hipótesis de fiabilidad es 0,70, aunque no se trate de un estándar absoluto, y los valores por debajo de 0,70 se han demostrado aceptables si la investigación tiene naturaleza exploratoria.

Debemos tener en cuenta, sin embargo, que la fiabilidad no asegura **validez**. La validez es la medida en que los indicadores miden «con precisión» lo que se supone que están midiendo. Por ejemplo, pueden ser fiables diversas medidas de cómo y por qué los consumidores compran productos, pero el investigador puede asumir erróneamente que miden la lealtad a la marca cuando de hecho son indicadores de intenciones de compra. En este caso, los indicadores son un conjunto fiable de medidas pero una medida inválida de lealtad a la marca. El supuesto de validez descansa en la especificación del investigador de indicadores para un constructo latente. Los medios para evaluar la validez en sus diversas formas se revisa en Bollen [22].

La fiabilidad y la varianza extraída (véase la sección siguiente) de un constructo latente deben calcularse por separado para cada constructo con indicadores múltiples en el modelo. Aunque LISREL y otros programas no los calculan directamente, se proporciona toda la información necesaria. La fiabilidad compuesta de un constructo se calcula como:

$$\text{Fiabilidad de constructo} = \frac{\left(\Sigma \text{ ponderaciones estandarizadas}\right)^2}{\left(\Sigma \text{ ponderaciones estandarizadas}\right)^2 + \Sigma\varepsilon_j}$$

donde las ponderaciones estandarizadas se obtienen directamente como resultado del programa, y ε_j es el error de medida de cada indicador [43]. El error de medida es 1,0 menos la fiabilidad del indicador, que es el cuadrado de la ponderación estandarizada del indicador. Las fiabilidades del indicador deberían exceder de 0,50, aunque corresponda aproximadamente a una ponderación estandarizada de 0,7.

Varianza extraída Otra medida de fiabilidad es la **medida de varianza extraída**. Esta medida refleja la cantidad total de la varianza de los indicadores tenida en cuenta por el constructo latente. Los mayores valores de la varianza extraída se producen cuando los indicadores son verdaderamente representativos del constructo latente. La varianza extraída es una medida complementaria del valor de la fiabilidad del constructo. La medida de varianza extraída se calcula como:

$$\text{Varianza extraída} = \frac{\left(\Sigma \text{ ponderaciones estandarizadas}^2\right)}{\left(\Sigma \text{ ponderaciones estandarizadas}\right)^2 + \Sigma\varepsilon_j}$$

Esta medida es muy parecida a la medida de la fiabilidad pero difiere en que las ponderaciones estandarizadas se elevan al cuadrado antes de sumarlas [43]. En general, se sugiere que el valor de la varianza extraída debería exceder de 0,50 para un constructo. Los ejemplos reales de los cálculos tanto para la medida de la fiabilidad como la varianza extraída, se ofrecen en el ejemplo del análisis factorial confirmatorio de este mismo capítulo.

Ajuste del modelo estructural

El examen más obvio del modelo estructural consiste en la significación de los coeficientes estimados. Los métodos de modelización de ecuaciones estructurales ofrecen no sólo coeficientes estimados sino también errores estándar y valores *t* calculados para cada coeficiente. Si podemos especificar el nivel de significación que juzgamos apropiado (es decir 0,05), entonces cada coeficiente estimado puede ser contrastado por la significación estadística (es decir, que sea diferente de cero) para las relaciones causales supuestas. Sin embargo, dadas las propiedades estadísticas del MLE y sus características (tamaños de muestra reducidos), se anima al investigador a ser conservador en la especificación del nivel de significación, eligiendo niveles pequeños (0,025 o 0,01) en lugar del tradicional nivel de 0,05.

La selección de un valor crítico también depende de la justificación teórica de las relaciones propuestas. Si se supone una relación positiva o negativa, se puede emplear un test de significación con una sola cola. Sin embargo, si el investigador no puede preespecificar la dirección de la relación, entonces se necesita utilizar un test de dos colas. La diferencia está en los valores críticos de la *t* utilizados para evaluar la significación. Por ejemplo, para el nivel de significación de 0,05, el valor crítico para un test de una sola cola es de 1,645, pero aumenta a 1,96 para un test de dos colas. Por tanto, el investigador puede detectar con mayor precisión las diferencias si puede utilizarse una teoría más fuerte en la especificación del modelo.

Como medida de la ecuación estructural, se calcula un coeficiente conjunto de determinación (R^2), similar al que se encuentra en la regresión múltiple. Aunque no pueda llevarse a cabo ningún test de significación estadística, ofrece una medida de ajuste relativa para cada ecuación estructural.

Los resultados de SEM pueden verse afectados por la multicolinealidad, al igual que en la regresión. Aquí el investigador debe tener cuidado de las correlaciones entre las estimaciones de los constructos en los resultados de SEM. Si aparece un valor muy elevado, entonces debe llevarse a cabo una acción correctiva. Esta acción puede incluir la destrucción de un constructo o la reformulación de relaciones causales. Aunque no se ha fijado ningún límite que defina lo que se consideran como correlaciones elevadas, los valores que exceden 0,90 siempre deberían ser examinados, y muchas veces las correlaciones que exceden 0,80 pueden ser indicadoras de problemas.

Comparación de modelos anidados o rivales

Las estrategias de modelización más habituales —estrategia de desarrollo de modelo o modelos rivales— consisten en la comparación de los resultados del modelo para determinar el modelo mejor ajustado de un conjunto de modelos. En una estrategia de desarrollo del modelo, el investigador comienza con un modelo inicial y sigue con una serie de reespecificaciones del modelo, con las que cada vez se espera mejorar el ajuste del modelo mientras se mantenga la concordancia con la teoría subyacente.

Para ayudar a la comparación entre modelos, se han desarrollado una gran cantidad de medidas para evaluar el ajuste del modelo. Una clase de medidas evalúa el ajuste del modelo conjunto en términos absolutos, ofreciendo una medida específica del ajuste. Una conclusión sobre estas medidas es que no tienen en cuenta el número de relaciones utilizadas en la obtención del ajuste del modelo. Para medir la parsimonia del modelo, se han propuesto una serie de medidas de ajuste de parsimonia. Su objetivo es determinado por el «ajuste por coeficiente», dado que el ajuste absoluto siempre mejorará a medida que se añaden coeficientes.

Anderson y Gerbing [4] propusieron un procedimiento exhaustivo para este propósito, en el que se especifican una serie de modelos rivales. Puede mostrarse que las diferencias entre modelos son simplemente la diferencia entre los valores de la chi-cuadrado para los diferentes modelos. A continuación se puede contrastar la significación de la diferencia de la chi-cuadrado con los grados de libertad apropiados, esto es, la diferencia en el número de coeficientes estimados para los dos modelos. El único requisito es que el número de constructos o indicadores sea el mismo, de tal forma que el modelo nulo sea el mismo para ambos modelos (es decir, son modelos anidados). El efecto de añadir o destruir una o más relaciones causales se puede contrastar también de esta forma haciendo comparaciones entre modelos con y sin las relaciones. Si los modelos no se hacen anidados (tener un número diferente de indicadores o constructos), el investigador debe basarse en las medidas de ajuste de parsimonia descritas previamente, en la medida en que el test de la chi-cuadrado no es apropiado para este caso.

Séptimo paso: Interpretación y modificación del modelo

Una vez que el modelo se considera aceptable, el investigador debería examinar en primer lugar los resultados y su correspondencia con la teoría propuesta. ¿Están corroboradas y son estadísticamente significativas las principales relaciones de la teoría? ¿Añaden los modelos rivales mayor perspectiva sobre las formulaciones alternativas de la teoría como para que puedan ser tenidas en cuenta? ¿Están todas las relaciones en la dirección supuesta (positiva o negativa)? Todas estas cuestiones y muchas más pueden ser contestadas a partir de los resultados empíricos. En el proceso de respuesta a estas tres cuestiones, el investigador puede tener necesidad de considerar dos supuestos de interpretación: el uso de las soluciones estandarizadas frente a las no estandarizadas y la reespecificación del modelo.

Soluciones estandarizadas frente a las no estandarizadas

Un aspecto de la evaluación de una relación estimada es la valoración del tamaño del parámetro. Al igual que en otras técnicas multivariantes, tales como la regresión múltiple, existe una diferencia notable en las soluciones estandarizadas y no estandarizadas en términos de su interpretación y uso. En los modelos de ecuaciones estructurales, los coeficientes estandarizados tienen todos igual varianza y un valor máximo de 1,0, por tanto, efectos tamaño muy aproximados, como se mostró con las ponderaciones beta en la regresión. Los coeficientes cercanos a cero tienen poco efecto sustantivo, si es que tienen alguno, mientras que un aumento en el valor corresponde a un aumento de la importancia en las relaciones causales. Los coeficientes estandarizados son útiles en la determinación de la importancia relativa, pero son específicos de una muestra y no son comparables entre las muestras.

Los coeficientes sin estandarizar corresponden a las ponderaciones de la regresión en una regresión múltiple en la que se expresan en términos de la escala del constructo, en este caso su varianza. Esto hace comparables a estos coeficientes para las muestras y retiene sus efectos escala. Dado que la escala varía para cada constructo, sin embargo, la comparación entre coeficientes es más difícil que con los coeficientes estandarizados.

Reespecificación del modelo

Una vez que la interpretación del modelo se ha completado, el investigador normalmente busca métodos para mejorar el ajuste del modelo y/o su correspondencia con la teoría subyacente. En tales casos, puede iniciar la **reespecificación del modelo**, el proceso de añadir o eliminar los parámetros estimados del modelo original. Antes de tratar algunos enfoques para identificar la modificación del modelo, aconsejamos al investigador hacer tales modificaciones con cuidado y sólo después de obtener justificación teórica para lo que se considera empíricamente deseable. Las modificaciones del modelo original deberían hacerse sólo después de una consideración deliberada. Si se hacen las modificaciones, el modelo debería tener validación cruzada (es decir, estimado sobre un conjunto distinto de datos) antes de que el modelo modificado sea aceptado.

Un proceso de reespecificación del modelo Antes de identificar cualquier posible reespecificación del modelo, el investigador debería clasificar todas las relaciones (estimadas o no) en una de dos categorías: teóricas o empíricas [88]. Las relaciones teóricas son esenciales a la teoría subyacente y no pueden ser modificadas. Están fuera de los límites de la reespecificación. La categoría empírica contiene relaciones que se añaden para mejorar el ajuste del modelo. Estas pueden ser reespecificadas. El objetivo es determinar un conjunto de modelos teóricos anidados, donde el conjunto de modelos se contempla como un conjunto de diferentes niveles de parsimonia para la misma teoría subyacente. De esta forma, el modelo se convierte en una serie de modelos rivales que llevan a varios niveles de corroboración para la teoría.

Indicadores empíricos de posibles reespecificaciones ¿Dónde puede buscar el investigador la mejora de los modelos? La primera indicación viene del examen de los residuos de la matriz de las predicciones de la covarianza y correlación. Los residuos estandarizados (también denominados residuos normalizados) representan las diferencias entre la matriz de covarianza o correlación observada y la matriz de covarianza o correlación estimada. Con las últimas versiones de LISREL (versión 7 y superiores), se ha mejorado el cálculo de los residuos y el patrón de evaluación de los residuos «significativos» ha cambiado (el umbral anterior estaba en $\pm$ 2,0). Los valores residuales mayores que $\pm$ 2,58 se consideran ahora estadísticamente significativos al nivel de 0,05. Los residuos significativos indican un error de predicción sustancial para un par de indicadores (es decir, una de las covarianzas o correlaciones de los datos de entrada originales). Un re-

siduo estandarizado indica sólo que existe diferencia pero no arroja luz acerca de cómo puede ser reducida. El investigador debe identificar el remedio mediante la suma o modificación de las relaciones causales [31, 49].

Otro apoyo en la evaluación del ajuste de un modelo especificado consiste en los **índices de modificación**, que se calculan para cada relación no estimada. El valor del índice de modificación corresponde aproximadamente a la reducción en la chi-cuadrado que se produciría si el coeficiente fuera estimado. Un valor de 3,84 o superior sugiere que se obtiene una reducción estadísticamente significativa en la chi-cuadrado cuando se estima el coeficiente. Aunque los índices de modificación pueden ser útiles en la evaluación del impacto de modificaciones basadas teóricamente, el investigador nunca hace cambios de modelo basándose sólo en índices de modificación. Este enfoque ateórico es completamente contrario al «espíritu» de la técnica y debería ser evitado en todos los casos. La modificación del modelo debería tener una justificación teórica antes de ser considerada e incluso entonces el investigador debería ser escéptico acerca de los cambios [72].

Además del índice de modificación, LISREL también ofrece un parámetro de cambio esperado, que hace referencia a la magnitud y la dirección de cada parámetro fijo (no estimado). Este parámetro difiere del índice de modificación en que no indica el cambio en el ajuste del modelo conjunto (χ^2) y en su lugar representa el cambio en el valor del parámetro efectivo. EQS ofrece el multiplicador de Lagrange y el estadístico de Wald, que evalúa el efecto de liberar un conjunto de parámetros simultáneamente.

Una vez hechas las modificaciones en el modelo, el investigador debe volver al paso 4 del proceso de siete pasos y reevaluar los modelos modificados. Si se anticipan modificaciones amplias del modelo, los datos deberían dividirse en dos muestras, una de ellas ofreciendo la base para la estimación y modificación del modelo, y la otra validación para el modelo final.

Una recapitulación del proceso de siete pasos

La modelización de ecuaciones estructurales ofrece al investigador más flexibilidad que cualquier otro método multivariante discutido. Pero junto con esta flexibilidad viene el potencial uso inapropiado del modelo. Un asunto primordial en cualquier aplicación de SEM, es que debería tener una fundamentación teórica incondicional, apoyada por un test confirmatorio robusto para una serie de relaciones casuales. Sin embargo, cuando se aplica el método de forma «exploratoria», el investigador se enfrenta con la muy elevada probabilidad de no conseguir resultados mediante el «husmeo» o «pesca» de datos y las relaciones identificadas tendrán poca generalización mediante la simple extrapolación de las relaciones específicas de los datos de la muestra que se están estudiando.

Dos ilustraciones de modelización de ecuaciones estructurales

La modelización de ecuaciones estructurales puede tratar una amplia variedad de relaciones causales. Entre los tipos más comunes de análisis desarrollados están el análisis factorial confirmatorio y la estimación de una serie de ecuaciones estructurales. Ahora ilustraremos ambos análisis siguiendo el proceso de siete pasos de SEM. En primer lugar, se examina la solución de dos factores proporcionada por el análisis factorial desarrollado en el Capítulo 3, mediante un análisis factorial confirmatorio. A continuación se proponen una serie de relaciones estructurales desde una nueva base de datos que trata un estudio expandido de las características del proveedor que el cliente considera importante.

Análisis factorial confirmatorio

El análisis factorial, como se discutió en el Capítulo 3, hace referencia a las pautas de exploración de las relaciones entre varias variables. Estas pautas se representan por lo que se denominan componentes principales o, más comúnmente, factores. Como las variables ponderan fuertemente sobre un factor, se hacen descriptores de las dimensiones subyacentes. Sólo con el examen de las ponderaciones de las variables sobre los factores se identifica el carácter de la dimensión subyacente.

En este punto, el lector puede ver ahora la similitud entre los objetivos del análisis factorial y el modelo de medida del SEM. Los factores son, en términos de los modelos medidos, las variables latentes. Cada variable actúa como indicador de cada factor (dado que cada variable tiene una ponderación para cada factor). Utilizado de esta forma, el análisis factorial es primordialmente una técnica exploratoria porque el investigador tiene un control limitado sobre las variables consideradas indicadores del constructo latente (es decir, aquellas variables «ponderadas» sobre cada factor). SEM, sin embargo, puede jugar un papel confirmatorio porque el investigador tiene un control completo sobre la especificación de los indicadores de cada constructo. Además, SEM permite un test estadístico de calidad del ajuste para la solución confirmatoria del factor propuesta, que no es posible con los componentes principales o el análisis de factor. El análisis de factor confirmatorio (CFA) es particularmente útil en la validación de las escalas para la medida de los constructos específicos [92].

Primer paso: Desarrollo de un modelo basado en la teoría

El uso confirmatorio de SEM se puede ilustrar mediante una síntesis de los ejemplos del análisis de componentes principales y del análisis factorial común desarrollados en el Capítulo 3. Como repaso, seis variables medían las percepciones de los encuestados de HATCO sobre las características de X_1, velocidad de entrega; X_2, nivel de precios; X_3, flexibilidad de precios; X_4, imagen del productor; X_6, imagen de la fuerza de ventas; y X_7, calidad de producto. La variable X_5, servicio conjunto, se omitió en el análisis factorial y será eliminada en este ejemplo también para mantener la comparabilidad.

El análisis factorial del Capítulo 3 indicó la existencia de dos dimensiones (factores). En primer lugar, los principales indicadores de una dimensión fueron las cuatro medidas de objetivo (velocidad de entrega, nivel de precio, flexibilidad de precio y calidad del producto). La segunda dimensión se caracterizaba por dos variables de imagen (imagen del productor y de la fuerza de ventas). Estas dos dimensiones de las percepciones del proveedor se caracterizan como acciones de estrategia específicas frente a evaluaciones más globales o afectivas. Por tanto, el modelo propuesto plantea dos factores (*estrategia* e *imagen*), con cada conjunto de variables actuando como indicadores de constructos separados. No existe razón para esperar percepciones no correlacionadas; por consiguiente, también se podrán correlacionar los factores.

Una serie de investigaciones recientes han demostrado los beneficios del uso del análisis factorial como un complemento a la teoría en la especificación de las ponderaciones de factor apropiado en el modelo de medida [48]. También se puede encontrar una discusión más detallada del procedimiento de estimación subyacente utilizando LISREL para la estimación de los componentes principales [37].

Segundo paso: Construcción de un diagrama de relaciones causales

El siguiente paso es representar las relaciones en un diagrama de secuencias. En este caso, los dos factores propuestos se consideran constructos exógenos. El diagrama de secuencias, incluyendo las variables que miden cada constructo, se muestran en la Figura 11.7. La correlación entre las percepciones se representa mediante la línea curva que conecta los dos constructos.

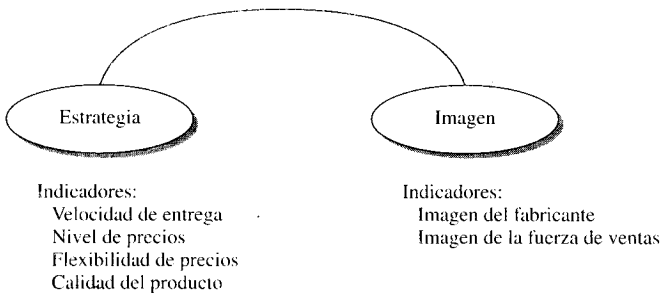


FIGURA 11.7. Diagrama de secuencia para el análisis factorial confirmatorio.

Tercer paso: Conversión de un diagrama de relaciones en un conjunto de modelos de ecuaciones estructurales y de medida

Dado que todos los constructos del diagrama de relaciones son exógenos, sólo necesitamos considerar el modelo de medida y las matrices de correlación asociadas para los constructos exógenos y los indicadores. Sin modelo estructural, el modelo de medida constituye todo el esfuerzo de modelización de ecuaciones estructurales (aquí hacemos referencia al análisis factorial *confirmatorio*).

TABLA 11.2. Modelo de medida de dos constructos

	Ponderaciones de indicadores como constructos	
Variables	*Estrategia*	*Imagen*
X_1 velocidad de entrega	L_1	
X_2 Nivel de precios	L_2	
X_3 Flexibilidad de precios	L_3	
X_4 Imagen del productor		L_4
X_6 Imagen de la fuerza de ventas		L_5
X_7 Calidad del producto	L_6	

TABLA 11.3. Notación LISREL para el modelo de medida

Indicador exógeno		*Constructo exógeno*		*Error*
X_1	=	$\lambda^x_{11}\xi_1$	+	δ_1
X_2	=	$\lambda^x_{21}\xi_1$	+	δ_2
X_3	=	$\lambda^x_{31}\xi_1$	+	δ_3
X_4	=	$\lambda^x_{42}\xi_2$	+	δ_4
X_5	=	$\lambda^x_{52}\xi_2$	+	δ_5
X_6	=	$\lambda^x_{61}\xi_1$	+	δ_6

Correlación entre constructos exógenos (ϕ)

	ξ_1	ξ_2
ξ_1	—	
ξ_2	ϕ_{21}	—

El modelo de medida se puede representar sencillamente por un modelo de dos constructos (estrategia e imagen), tal y como se muestra en la Tabla 11.2. Además, se supone que los dos constructos están correlacionados, y no se proponen mediciones correlacionadas dentro del constructo.

Para el lector interesado, la notación apropiada de LISREL (que se especifica en el Apéndice 11.A) es como la que se muestra en la Tabla 11.3.

Cuarto paso: Elección del tipo de matriz de entrada y estimación del modelo propuesto

Entrada de datos

Con la modelización de ecuaciones estructurales, se obtendrá tanto la matriz de covarianza como de correlación. A efectos de análisis factorial confirmatorio, se puede emplear cualquiera de las matrices de entrada. Sin embargo, dado que el objetivo es una exploración de la pauta de interrelaciones, las correlaciones son el tipo de datos de entrada preferidos. La matriz de correlación de las seis variables se muestra en la Tabla 11.4.

Los supuestos básicos de la modelización estructural, similares a otros métodos multivariantes, se examinaron en el Capítulo 2. A efectos ilustrativos, se considera que las variables cumplen estos supuestos, dirigiendo nuestra atención a los resultados de las estimaciones infractoras.

TABLA 11.4. Matriz de correlación para el análisis factorial confirmatorio

Variables	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	1,000					
X_2 Nivel de precios	–0,349	1,000				
X_3 Flexibilidad de precios	0,509	–0,487	1,000			
X_4 Imagen del fabricante	0,050	0,272	–0,116	1,000		
X_6 Imagen del personal de ventas	0,077	0,186	–0,034	0,788	1,000	
X_7 Calidad del producto	–0,483	0,470	–0,448	0,200	0,177	1,000

Estimación del modelo

Para la estimación del modelo de medida y las correlaciones del constructo se utiliza el LISREL [64, 67]. Consideremos la estimación del modelo de medida para los constructos con más de una variable: debido al procedimiento de estimación, el constructo debe hacerse «invariante a la escala», lo que significa que los indicadores de los constructos deben ser «estandarizadas» de forma que se hagan los constructos comparables [65, 71, 81]. Existen dos enfoques comunes para este procedimiento. En primer lugar, una de las ponderaciones de cada constructo puede anclarse al valor fijo 1,0. El segundo enfoque es estimar la varianza del constructo directamente. Con cada enfoque se obtienen las mismas estimaciones; pero a efectos de contraste de la teoría, se recomienda el segundo enfoque (estimación de la varianza del constructo) [4]. En este ejemplo, se emplea la segunda aproximación. El conjunto completo de comandos utilizados por el análisis LISREL se muestra en el Apéndice A.

Quinto paso: Valoración de la identificación del modelo estructural

La identificación es un asunto relativamente simple en el análisis factorial confirmatorio y los procedimientos de diagnóstico del programa son suficientes para detectar los problemas de identifi-

cación. El problema más común que puede producirse es que se suponga que múltiples variables sean indicadores de dos o más constructos. Es posible que los problemas de identificación se produzcan en muchos casos, pero no se encuentran en la mayoría de los análisis factoriales confirmatorios, y las posibilidades de que ocurran se minimizan por el uso de fuertes fundamentos teóricos de especificación del modelo de medida.

Sexto paso: Evaluación de criterios de calidad del ajuste

Estimaciones infractoras

La Tabla 11.5 contiene las estimaciones de LISREL del modelo de medida y de las correlaciones de los constructos. La ponderación de la imagen del fabricante (X_4) es mayor que 1,0. Se trata, por tanto, de una estimación infractora. El lector debería notar asimismo un correspondiente valor de medida de error negativo (conocido como caso Heywood) para la misma variable. Tales estimaciones son teóricamente inapropiadas y deben corregirse antes de que el modelo sea interpretado y la calidad del ajuste evaluada.

TABLA 11.5. Resultados del análisis factorial confirmatorio: Modelo inicial

PONDERACIONES DE LOS CONSTRUCTOS

	Constructo exógeno	
Variables	*Estrategia*	*Imagen*
X_1 Velocidad de entrega	0,644	0,000
X_2 Nivel de precios	-0,654	0,000
X_3 Flexibilidad de precios	0,722	0,000
X_4 Imagen del fabricante	0,000	1,151
X_6 Imagen del personal de ventas	0,000	0,685
X_7 Calidad del producto	0,689	0,000

CORRELACIONES ENTRE CONSTRUCTOS (FACTORES) LATENTES

	Estrategia	*Imagen*
Estrategia	1,000	
Imagen	–0,175	1,000

ERROR DE MEDIDA PARA INDICADORES

Variables	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	0,585					
X_2 Nivel de precios	0,000	0,573				
X_3 Flexibilidad de precios	0,000	0,000	0,479			
X_4 Imagen del fabricante	0,000	0,000	0,000	–0,325		
X_6 Imagen del personal de ventas	0,000	0,000	0,000	0,000	0,531	
X_7 Calidad del producto	0,000	0,000	0,000	0,000	0,000	0,526

Reespecificación del modelo

Existen varios remedios para corregir la estimación infractora, incluyendo la eliminación de la variable infractora. En esta situación, la variable será retenida y la correspondiente varianza del error (–0,325 sobre la diagonal de la matriz de medida del error) se fijará a un valor reducido (0,005) para asegurar que la ponderación será ahora menor que 1,0. A continuación se reestima el modelo.

Los resultados del modelo reespecificado se muestran en la Tabla 11.6. Al examinar los resultados, no hay ninguna estimación infractora; por tanto podemos proceder a evaluar la calidad del ajuste del análisis factorial confirmatorio.

TABLA 11.6. Resultados del análisis factorial confirmatorio: Modelo revisado

PONDERACIONES DEL CONSTRUCTO (VALORES DE LA *t* ENTRE PARÉNTESIS)

	Constructo exógeno	
Variables	*Estrategia*	*Imagen*
X_1 Velocidad de entrega	0,643 (6,263)	0,000
X_2 Nivel de precios	–0,654 (–6,381)	0,000
X_3 Flexibilidad de precios	0,718 (7,118)	0,000
X_4 Imagen del fabricante	0,000	0,997 (14,001)
X_6 Imagen del personal de ventas	0,000	0,790 (9,467)
X_7 Calidad del producto	–0,692 (–6,821)	0,000

CORRELACIONES ENTRE CONSTRUCTOS LATENTES (VALORES DE LA *t* ENTRE PARÉNTESIS)

	Estrategia	*Imagen*
Estrategia	1,000	
Imagen	–0,202 (–1,826)	1,000

ERROR DE MEDIDA ENTRE LOS INDICADORES

Variables	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	0,586					
X_2 Nivel de precios	0,000	0,573				
X_3 Flexibilidad de precios	0,000	0,000	0,484			
X_4 Imagen del fabricante	0,000	0,000	0,000	0,005		
X_6 Imagen del personal de ventas	0,000	0,000	0,000	0,000	0,376	
X_7 Calidad del producto	0,000	0,000	0,000	0,000	0,000	0,521

Ajuste conjunto del modelo: Modelo revisado

La primera evaluación del ajuste del modelo debe hacerse para el modelo conjunto. En el análisis factorial confirmatorio, el ajuste del modelo conjunto representa el grado en que los indicadores especificados representan los constructos supuestos. La última versión de LISREL (versión 8) pro-

porciona ahora un amplio rango de calidad del ajuste, tal y como se describe en el Apéndice 11C. A efectos del análisis factorial confirmatorio, nos centramos sólo en un número limitado de medidas de cada tipo. Una discusión más detallada de todas las medidas la encontramos en el segundo ejemplo de estimación de un modelo estructural completo y evaluación de los modelos rivales. Los tres tipos de medidas de ajuste conjunto utilizadas en el análisis pueden representarse por las medidas de ajuste parsimonioso, incremental y absoluto.

Medidas de ajuste absoluto LISREL ofrece medidas de calidad de ajuste, como se muestra en la Tabla 11.7. La primera medida es el ratio de verosimilitud del estadístico chi-cuadrado. El valor (χ^2 = 15,745, nueve grados de libertad) tiene un nivel de significación estadística de 0,072, por encima del nivel mínimo de 0,05, pero no por encima del niveles más conservadores de 0,10 ó 0,20. Este estadístico muestra cierta evidencia para pensar que las diferencias de las matrices efectivas y previstas no son significativas, indicativo de un ajuste aceptable. Además, el tamaño muestral de 100 está dentro del rango aceptable de aplicación de esta medida. También cumple los criterios de cinco observaciones por parámetro estimado. Nótese sin embargo, que existe potencial para el χ^2 con tamaños de muestra de 100 o menores para poner de relieve que no hay diferencias incluso cuando el modelo no tenga relaciones significativas, y siempre será aconsejable en situaciones como ésta aumentar el tamaño muestral. Por tanto, deben emplearse medidas de ajuste adicionales. El índice de bondad del ajuste (GFI) tiene un valor de 0,949, que es bastante alto, pero no se ajusta a la parsimonia del modelo. Finalmente, la raíz cuadrada de la media del cuadrado de los residuos (RMSR) indica que la correlación residual media es 0,075, considerada aceptable dadas las fuertes correlaciones de la matriz de correlación original*. Incluso aunque todas las medidas caigan dentro de unos límites aceptables, son necesarios los índices de ajuste incremental y de parsimonia para asegurar la aceptabilidad del modelo desde otras perspectivas.

Medidas de ajuste incremental El siguiente tipo de medidas de bondad del ajuste evalúan el ajuste incremental del modelo comparado con un **modelo nulo**. En este caso, el modelo nulo se supone como un modelo de un único factor sin error de medida. Los resultados de estimación para el modelo nulo también se muestran en la Tabla 11.7. (Las tarjetas de control de la estimación del modelo nulo se muestran en el Apéndice A.)

TABLA 11.7. LISREL medidas de bondad del ajuste para CFA: Modelos revisados y nulos

Modelo revisado	
Chi-cuadrado (χ^2)	15,745
Grados de libertad[a]	9
Nivel de significación	0,072
Índice de bondad del ajuste	0,949
Índice de bondad del ajuste adaptado (AGFI)	0,881
Residuo cuadrático medio (RMSR)	0,075
Modelo nulo	
Chi-cuadrado (χ^2)	212,033
Grados de libertad	15
Nivel de significación	0,000

[a] El número esperado de grados de libertad para el modelo revisado es 8 [(0,5)(6)(7) – 13]. Sin embargo, dado que el error de medida de X_4 se fijó a 0,005 y no se estimó, se dispone de un grado de libertad adicional.

* En la bibliografía consultada sobre modelos estructurales publicada en España, no se suelen traducir estas medidas, que se conocen como *Goodness-of-fit* (GFI) y *root mean square residual* (RMSR). Hemos preferido traducir estos términos para clarificar el capítulo a aquellos lectores que no conocen estas expresiones. (N. de la R.T.)

El modelo nulo tiene un valor χ^2 de 212,033 con 15 grados de libertad. Con esta información, podemos calcular ahora dos medidas de ajuste incremental, el índice de Tucker-Lewis y el índice de ajuste normado.

Índice Tucker-Lewis (TLI)

$$\text{TLI} = \frac{(\chi^2_{\text{null}} / gl_{\text{null}}) - (\chi^2_{\text{propuesta}} / gl_{\text{propuesta}})}{(\chi^2_{\text{null}} / gl_{\text{null}}) - 1} =$$

$$= \frac{(212{,}033/15) - (15{,}75/9)}{(212{,}033/15) - 1} = 0{,}943$$

Índice del ajuste normado (NFI)

$$\text{NFI} = \frac{\chi^2_{\text{null}} - \chi^2_{\text{propuesta}}}{\chi^2_{\text{null}}} =$$

$$= \frac{(212{,}033 - 15{,}75)}{212{,}033} = 0{,}926$$

ambas medidas del ajuste incremental exceden el nivel recomendado de 0,90, aportando evidencia suplementaria para aceptar el modelo propuesto.

Medidas de ajuste de la parsimonia Las últimas medidas del modelo conjunto evalúan la parsimonia del modelo propuesto mediante la evaluación del ajuste del modelo frente al número de coeficientes estimados (o a la inversa, los grados de libertad) necesarios para conseguir ese nivel de ajuste. Dos medidas apropiadas para la evaluaciones directas del modelo son el índice ajustado de bondad del ajuste (AGFI) y la chi-cuadrado normada. La AGFI se ofrece en el programa LISREL. El valor AGFI de 0,881 está cerca del nivel recomendado de 0,90; por tanto, se puede aceptar con precauciones esta medida. La chi-cuadrado normada (χ^2/gl) tiene un valor de 1,75 (15,75/9). Este valor cae dentro de los niveles recomendados de 1,0 a 2,0. Combinado con el AGFI, este resultado permite la evidencia condicional de la parsimonia del modelo.

En resumen, las diversas medidas globales de calidad del ajuste aportan suficientes evidencias para considerar los resultados como una aceptable representación de los constructos supuestos.

Ajuste del modelo de medida

Una vez que se ha aceptado el modelo conjunto, se pueden evaluar por separado cada uno de los constructos mediante (1) el examen de las ponderaciones de los indicadores de significación estadística y (2) la evaluación de la fiabilidad del constructo, así como la extracción de la varianza. Primero, para cada variable, los valores *t* asociados con cada una de las ponderaciones exceden a los valores críticos para el nivel de significación 0.05 (valor crítico = 1,96) y el nivel de significación de 0,01 (valor crítico = 2,576). Por tanto, todas las variables están significativamente relacionadas con sus constructos específicos, verificando las relaciones propuestas entre los indicadores y los constructos.

Posteriormente, debemos estimar las fiabilidades y las medidas de varianza extraída para cada constructo con el fin de evaluar si los indicadores especificados son suficientes en su representación de los constructos. Los cálculos de cada medida se muestran en la Tabla 11.8. Ambos constructos (0,772 y 0,893) exceden el nivel recomendado de 0,70, aunque debemos recordar que la fiabilidad del constructo 2 (imagen) está inflada como resultados de la reespecificación del error de medida para la variable X_4 con un valor casi de cero, para eliminar los valores de varianza de error negativos.

TABLA 11.8. Fiabilidad y estimaciones de varianza extraída para constructos en el análisis factorial confirmatorio

Fiabilidad

$$\text{Fiabilidad de constructo} = \frac{(\text{Suma de las ponderaciones estandarizadas})^2}{(\text{Suma de las ponderaciones estandarizadas})^2 + (\text{Suma de error de medida del indicador})}$$

Suma de las ponderaciones estandarizadas[a]

Estrategia = 0,643 + 0,654 + 0,718 + 0,692 = 2,707

Imagen = 0,997 + 0,790 = 1,787

Suma del error de medida[b]

Estrategia = 0,586 + 0,573 + 0,484 + 0,521 = 2,164

Imagen = 0,005 +0,376 = 0,381

Cálculo de fiabilidad

$$\text{Estrategia} = \frac{(2{,}707)^2}{(2{,}707)^2 + 2{,}164} = 0{,}772$$

$$\text{Imagen} = \frac{(1{,}782)^2}{(1{,}787)^2 + 0{,}381} = 0{,}893$$

Varianza extraída

$$\text{Varianza extraída} = \frac{\text{Suma de los cuadrados de las ponderaciones estandarizadas}}{\text{Suma de los cuadrados de las ponderaciones estandarizadas + Suma de error de medida del indicado}}$$

Suma de los cuadrados de las ponderaciones estandarizadas

Estrategia = $0{,}643^2 + 0{,}654^2 + 0{,}718^2 + 0{,}692^2 = 1{,}836$

Imagen = $0{,}997^2 + 0{,}790^2 = 1{,}618$

Cálculo de la varianza extraída

$$\text{Estrategia} = \frac{1{,}836}{1{,}836 + 2{,}164} = 0{,}459$$

$$\text{Imagen} = \frac{1{,}618}{1{,}618 + 0{,}381} = 0{,}809$$

[a] A efectos de calcular la fiabilidad de un constructo, se pueden ignorar los signos de la ponderación. Sin embargo, si el investigador desea calcular una medida de suma (como una escala total) para los indicadores, las variables con ponderaciones negativas deben ser invertidas en magnitud. Esto asegura que los indicadores negativos no compensan los indicadores positivos.

[b] El error de medida del indicador puede calcularse como 1 – (ponderación estandarizada)^2 o la diagoanl de la matriz de correlación del error de medida (matriz teta-delta) en el resultado LISREL.

Para la medidas de varianza extraída, el constructo 1 (estrategia) tiene un valor de 0,459, que está relativamente cerca de un porcentaje recomendado del 50 por ciento. El constructo 2 (imagen), con un valor de 0,809, de nuevo excede sustancialmente el nivel recomendado. El menor nivel de varianza extraída para el constructo 1 indica que más de la mitad de la varianza de los indicadores especificados no se tiene en cuenta para el constructo. Este resultado puede llevar al investigador a explorar ponderaciones adicionales para estos indicadores sobre los otros constructos, si está teóricamente justificado.

Otro coeficiente estimado en el modelo de medida es la correlación entre los dos constructos. El valor *t* para la correlación obtenido de –0,202 es 1,826 (véase Tabla 11.6) que cae por debajo del valor crítico para el nivel de significación de 0,05, aunque sólo ligeramente. Esto indica que puede haber evidencia débil (sólo al nivel de significación del 0,10) para creer que los constructos están correlacionados.

Resumen

Los resultados de calidad del ajuste del modelo conjunto y las evaluaciones del modelo de medida arrojan evidencia sustancial para confirmar el modelo propuesto de dos factores. Aunque se han conseguido resultados aceptables con este modelo, el siguiente paso explora las modificaciones posibles que pueden mejorar los resultados del modelo, si existe una justificación teórica para ello.

Séptimo paso: Interpretación y modificación del modelo

Los objetivos del análisis factorial confirmatorio son (1) verificar la estructura de factor del modelo propuesto y (2) explorar si son necesarias modificaciones significativas. Al hacerlo, examinamos en primer lugar las ponderaciones de factor y las correlaciones de factor para buscar a continuación las reespecificaciones posibles del modelo.

Interpretación

La interpretación del análisis factorial confirmatorio es bastante similar a aquella considerada para el análisis factorial en el Capítulo 3. En nuestro ejemplo, las cuatro variables tienen ponderaciones significativas sobre el primer constructo y las ponderaciones de las otras dos variables, son significativas en el segundo constructo. Al comparar estos resultados con el análisis factorial de las misma variables del Capítulo 3, encontramos que las conclusiones son similares, pero también hay un par de diferencias. En primer lugar, los signos de las ponderaciones son los inversos de los encontrados en el Capítulo 3. Esto no modifica la interpretación del constructo, pero afecta a la relación del segundo constructo, como se muestra en la correlación entre constructos, que ahora es negativa (en contraposición a la correlación negativa que apareció en el Capítulo 3) porque las dos variables (X_1 y X_3) con ponderaciones negativas en el Capítulo 3 resultan ahora en ponderaciones positivas de CFA. En segundo lugar, la correlación entre los constructos aumenta, como se debe a que el proceso de estimación tiene en cuenta el error de medida de los dos factores, que aumentan los parámetros estimados. Recordemos que esto se produce porque si la correlación observada se calcula con error de medida (que debería reducirse de alguna forma), entonces la «verdadera» relación será mayor.

Reespecificación del modelo

Se pueden identificar las posibles modificaciones del modelo propuesto a través del examen de los residuos normalizados y los índices de modificación. Sin embargo, en cualquier caso, las modificaciones deben ser teóricas antes de que el modelo reespecificado pueda ser contrastado. En ese ejemplo, las únicas posibles modificaciones sería la posibilidad de que una de las variables actúe como indicador para ambos constructos.

Los residuos normalizados se muestran en la Tabla 11.9 y el examen revela que sólo un valor excede 2,58. Por tanto, sólo una correlación de la matriz de entrada original tiene un residuo estadísticamente significativo, cayendo dentro del rango aceptable (de que uno de los 20 residuos excediera el valor de 2,58 por casualidad). La Tabla 11.9 también contiene los índices de modificación del modelo de medida. Dos de los indicadores (velocidad de entrega y nivel de precio) tienen modificaciones de indicadores por encima del nivel sugerido (3,84) para una posible reespecificación del modelo. En este caso, los índices de modificación indican que estas variables pueden ser indicadores tanto del segundo constructo así como del primero (es decir, ponderaciones múltiples).

La evidencia teórica para tal estructura no puede encontrarse; por tanto, el modelo no está reespecificado. Si la reespecificación del modelo se basa sólo en los valores de los índices de modificación, el investigador está extrapolando la unicidad de estos datos particulares, y el resultado será probablemente ateórico pero estadísticamente significativo, de forma que el modelo tiene escasa generalizabilidad y un uso limitado en la constrastación de las relaciones causales.

TABLA 11.9. Residuos normalizados e índices de modificación del análisis factorial confirmatorio

RESIDUOS NORMALIZADOS

Variables	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	0,000					
X_2 Nivel de precios	1,826	0,000				
X_3 Flexibilidad de precios	1,535	–0,595	0,000			
X_4 Imagen del fabricante	2,707	2,157	0,508	0,000		
X_6 Imagen del personal de ventas	2,234	1,028	1,067	0,000	0,000	
X_7 Calidad del producto	1,091	0,523	1,959	1,011	0,868	0,000

ÍNDICES DE MODIFICACIÓN

	Ponderaciones de constructo	
Variables	*Estrategia*	*Imagen*
X_1 Velocidad de entrega	0,000	7,359
X_2 Nivel de precios	0,000	4,364
X_3 Flexibilidad de precios	0,000	0,269
X_4 Imagen del fabricante	0,321	0,000
X_6 Imagen del personal de ventas	0,321	0,000
X_7 Calidad del producto	0,000	1,029

	Indicador del error de medida					
Variables	X_1	X_2	X_3	X_4	X_6	X_7
X_1 Velocidad de entrega	0,000					
X_2 Nivel de precios	3,333	0,000				
X_3 Flexibilidad de precios	2,357	0,355	0,000			
X_4 Imagen del fabricante	1,712	2,182	0,216	0,321		
X_6 Imagen del personal de ventas	0,182	0,046	0,927	0,321	0,000	
X_7 Calidad del producto	1,191	0,274	3,837	0,038	1,009	0,000

Modelos de análisis factorial de orden superior

Se puede obtener una perspectiva adicional sobre la estructura analítica del factor con la introducción de modelos de factor de orden superior (también conocidos como de segundo orden) [62, 74]. El modelo de análisis factorial que acabamos de examinar se conoce como un *modelo factorial de primer orden*. En este tipo de modelo confirmatorio, el investigador especifica sólo un nivel de factores (el primer orden) que está correlacionado. Pero con ello se asume que los factores, aunque correlacionados, son constructos aislados. ¿Qué ocurre si el investigador tiene un constructo con varias facetas o dimensiones, tales como imagen de establecimiento o satisfacción con el trabajo? Sería necesario ver si están correlacionados, pero lo que se necesita realmente es un medio de demostrar las relaciones estructurales entre las facetas y las dimensiones, con lo que se permitiría un juicio más profundo para la dimensionalidad de los constructos de satisfacción con el trabajo e imagen del establecimiento.

Esto se realiza a través de la especificación de un *modelo factorial de segundo orden* que plantea que los factores de primer orden estimados son, en realidad, subdimensiones de un constructo más amplio; en este ejemplo, la satisfacción con el trabajo y la imagen del establecimiento. A efec-

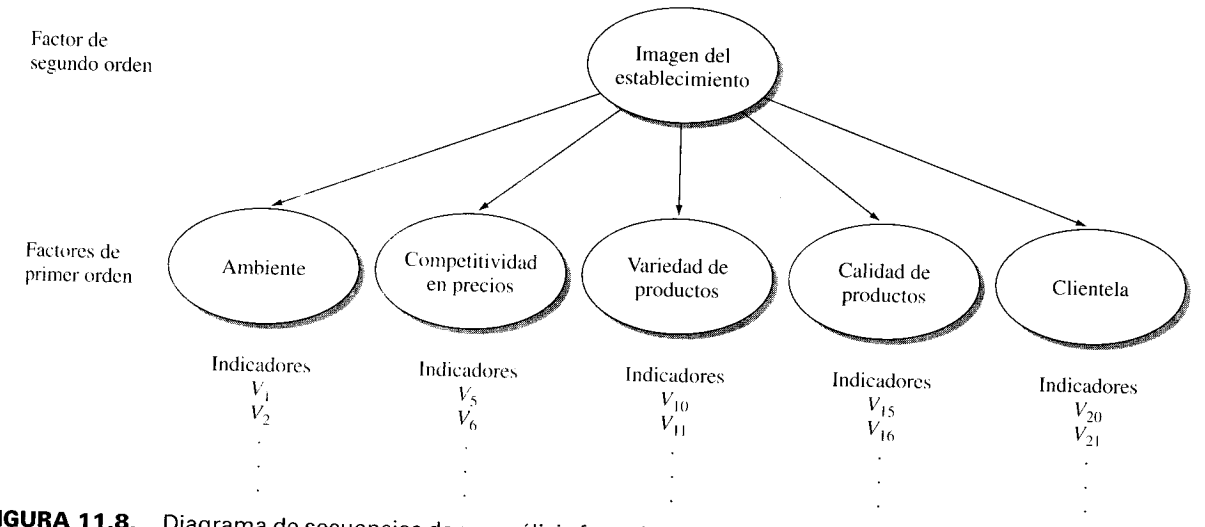


FIGURA 11.8. Diagrama de secuencias de un análisis factorial de segundo orden de la imagen del establecimiento.

tos ilustrativos, supone que la imagen del establecimiento se supone que contiene cinco dimensiones (ambiente, competitividad en precios, variedad de productos, calidad del producto y clientela). Siguiendo los procedimientos descritos previamente, se puede contrastar un modelo de cinco factores en el que cada uno de estos cinco constructos están relacionadas (o surgen de) un factor de segundo orden.

Existen dos características únicas del modelo de segundo orden. En primer lugar, los factores de segundo orden se convierten en factores exógenos, mientras que los factores de primer orden son endógenos. Por esto, decimos que el factor de segundo orden «causa» los factores de primer orden. Esta es la razón de por qué las flechas van del factor de segundo orden al factor de primer orden. Aunque podemos tener varios juicios que midan cada una de las cinco dimensiones, el segundo factor es completamente latente, no observable y no medible. Para muchos investigadores esto parece muy abstracto o como un enfoque de «caja negra». Pero debemos recordar que en realidad representa un concepto abstracto que da lugar a los constructos latentes. Los pasos básicos y los procedimientos para estimar un modelo de orden superior se pueden encontrar fácilmente en la literatura [48, 75]. Los procedimientos para contrastar modelos de factores de orden superior también están bien documentados [10].

Resumen

El análisis factorial confirmatorio ofrece un soporte adecuado al modelo propuesto basándose en los análisis factoriales desarrollados en el Capítulo 3. Sin embargo, debido a que el modelo tiene inconsistencias de los datos (el caso de Heywood para X_4, imagen del fabricante), el investigador debería obtener datos adicionales y contrastar el modelo con una nueva matriz de correlación para asegurar la generalización a múltiples muestras.

Estimación de un modelo de relaciones con SEM

El siguiente ejemplo de SEM examina un conjunto de dos relaciones causales, las percepciones de los clientes de HATCO en relación a su nivel de uso de los productos HATCO y su satisfacción con la empresa. El modelo estructural permite entender no sólo la importancia relativa de las diversas percepciones de HATCO en determinadas compras sino también las posibles fuentes de satisfacción de los clientes con HATCO.

Para evitar la duplicación del análisis factorial confirmatorio, ya examinado, se realiza un nuevo estudio entre los clientes de HATCO. Un total de 136 clientes calificaron a HATCO sobre un número de posibles determinantes de compra. Este estudio es similar a la base de datos originales introducida en el Capítulo 1 en el que se preguntó a los encuestados que calificaran a HATCO con una escala de 0 a 10 (0 = muy mal, y 10 = excelente), indicando con ello su satisfacción con la empresa. Las relaciones causales hacen referencia a la determinación de qué elementos de HATCO determinan el nivel de compras de los clientes y su satisfacción. Difiere en los atributos y en los tipos de relaciones examinadas.

Como el ejemplo anterior, la discusión sigue un proceso de siete pasos, que ilustran los supuestos y la interpretación de cada paso.

Primer paso: Desarrollo de un modelo basado en la teoría

El propósito de este análisis es entender mejor cómo las percepciones de los clientes de HATCO afectan a su comportamiento y actitudes. Como ya se ha expuesto, 136 clientes proporcionaron evaluaciones de 13 atributos para todo el espectro de modos en los cuales podían relacionarse con HATCO. Los atributos se obtuvieron mediante un proceso desarrollado en dos fases. En primer lugar, todos los jefes de zona participaron en una serie de grupos de trabajo en los que se discutieron estos temas, entre otros. De estas discusiones, se identificaron 27 atributos como posibles ítems evaluativos. Este conjunto de 27 ítems se presentó a los jefes de distrito, que calificaron cada ítem en términos de sus percepciones sobre la importancia de cada atributo. El análisis del conjunto de 27 ítems para ambas muestras resultó en un conjunto final de 13 ítems compuestos de tres áreas de evaluación. Las dimensiones y la evaluación de los ítems se muestran en la Tabla 11.10.

Una vez que se han establecido las dimensiones de evaluación que actúan como variables independientes, el último paso es determinar sus relaciones con las decisiones de compra y satisfacción. Dado que la satisfacción se basa en una evaluación del cliente de las experiencias pasadas con HATCO, la relación causal apropiada para los niveles de satisfacción actuales se predice por los niveles de compra a HATCO. Aunque se puede plantear una relación causal que represente a las compras surgiendo *de* la satisfacción, este método exigiría un enorme grado de control sobre los aspectos longitudinales de la recogida de datos necesaria para establecer el orden temporal. Sin embargo, sabemos que el nivel de compra refleja el comportamiento de compras pasado, de forma que puede ser un predictor válido de la satisfacción actual. El investigador debe examinar siempre cada relación propuesta desde una perspectiva teórica para asegurarse de que los resultados son

TABLA 11.10. Dimensiones de evaluación y atributos de la empresa para el nuevo estudio de HATCO

Dimensión de evaluación	*Atributo de la empresa*
Factores de empresa y producto	X_1 Calidad del producto
	X_2 Precisión de facturación
	X_3 Apoyo técnico
	X_4 Introducción de nuevos productos
	X_5 Fiabilidad de entrega
	X_6 Servicio al cliente
Factores de precio	X_7 Valor del producto
	X_8 Proveedor precios bajos
	X_9 Posición negociadora (políticas)
Factores de relación de compra	X_{10} Confluencia de intereses
	X_{11} Integridad y honestidad
	X_{12} Flexibilidad
	X_{13} Resolución de problemas

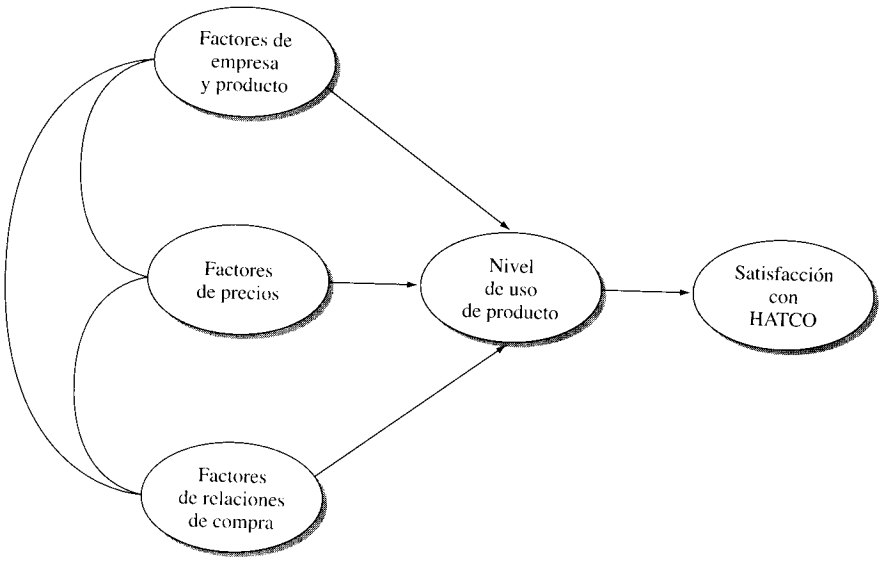


FIGURA 11.9. Diagrama de relaciones para SEM.

conceptualmente válidos. En este caso, el vínculo compra-satisfacción y las estimaciones hechas a partir de esta relación podrían revertirse rápidamente, porque la correlación opera igualmente bien en ambas relaciones. Pero los resultados podrían ser potencialmente engañosos porque existe un fallo conceptual en la relación inversa si se estima con estos datos.

Segundo paso: Construcción de un diagrama de relaciones causales

Una vez que se ha desarrollado un conjunto de relaciones causales, el siguiente paso consiste en representar las relaciones en un diagrama de relaciones (véase Figura 11.9). Como vemos, las tres dimensiones de evaluación actúan como variables exógenas, cada una de ellas relacionadas con el nivel de uso de producto, que es una de las dos variables endógenas. A continuación se plantea el uso de producto como el único predictor de la satisfacción. Las dimensiones de evaluación también podrían ser fuentes de satisfacción; sin embargo, estas relaciones causales no se proponen inicialmente, sino que en su lugar se exploran a través de la contrastación de especificaciones de modelos alternativas (es decir, modelos rivales). El diagrama de secuencia también indica que se propone que las tres dimensiones exógenas estén interconectadas. Aunque se supone que las dimensiones de evaluación son distintas, se reconoce que ciertas percepciones están compartidas y que, por tanto, existen correlaciones entre los constructos.

Tercer paso: Conversión del diagrama de relaciones en un conjunto de ecuaciones estructurales y especificación del modelo de medida

El diagrama de relaciones nos da los fundamentos para la especificación de las ecuaciones estructurales y las correlaciones propuestas (1) entre los constructos exógenos y (2) entre las ecuaciones estructurales. A partir del modelo de relaciones, el investigador puede construir una serie de ecuaciones estructurales (una para cada constructo endógeno) que constituya el modelo estructural. A continuación, se especifica el modelo de medida donde se asignan los indicadores a cada constructo (exógenos y endógenos).

Especificación de las ecuaciones estructurales

La especificación de una ecuación estructural para cada constructo exógeno (nivel de uso y satisfacción) debe especificar las relaciones tanto de los constructos exógenos como de los endógenos. Los cuatro coeficientes a estimar en las ecuaciones estructurales pueden expresarse como sigue:

	Exógenos			*Endógenos*	
Variable endógena	*Factores de empresa y producto*	*Factores basados en el precio*	*Factores relaciones de compra*	*Factores de Nivel de uso*	*Nivel de satisfacción*
Nivel de uso	b_1	b_2	b_3		
Satisfacción				b_4	

Nótese que los términos de error de cada ecuación se omitieron, pero deben ser incluidos cuando se estime el modelo.

Especificación del modelo de medida

El modelo de medida especifica la correspondencia de indicadores para los constructos. Para las dos variables endógenas, sólo se dispone de indicadores aislados. Por tanto, el investigador tiene que especificar su fiabilidad. Aunque se pueden hacer estimaciones del nivel error de medida, a efectos ilustrativos el error de medida se fijará en cero (indicando la medición perfecta de ambos constructos). La relajación de estos supuestos proporciona las bases para un modelo alternativo. (Este modelo alternativo no está contrastado en nuestra discusión, pero se anima al lector a considerar el análisis y examinar los efectos de los diversos niveles de medida del error.)

Los constructos exógenos se miden por los 13 indicadores que reflejan las dimensiones descritas anteriormente. No hacemos aquí una especificación formal del modelo de medida para los constructos exógenos, y el lector puede referirse a los ejemplos anteriores para detalles más específicos sobre este proceso si fuese necesario. Tampoco existen casos para los que los indicadores debieran estar correlacionados; por tanto, no se permiten correlaciones del error de medida, tanto en el modelo inicial o como base para las modificaciones del modelo.

Correlaciones entre constructos e indicadores

Además de las ecuaciones estructurales, las correlaciones han de ser especificadas entre el conjunto de constructos exógenos, así como entre cualquier relación estructural. En este ejemplo, cada constructo exógeno (factores de empresa y producto, factores basados en el precio y factores relacionados con la compra) se permite que esté correlacionado con los otros dos constructos exógenos. Esto representa la realización de influencias compartidas sobre estos constructos externos que no están especificados en el modelo, provocando por tanto su correlación. No existen correlaciones, sin embargo, entre las ecuaciones estructurales.

Cuarto paso: Elección del tipo de matriz de entrada y estimación del modelo propuesto

Con el modelo completamente especificado, los siguientes pasos son contrastar los datos para ver si se cumplen las hipótesis subyacentes al modelo de ecuaciones estructurales, seleccionar el tipo

de matriz de datos de entrada (covarianzas o correlaciones) a utilizar en la estimación del modelo y estimar los modelos de medida y estructurales.

Entrada de datos

Aunque el diseño del estudio asegura la independencia y aleatoriedad de las respuestas, todavía hay que evaluar las características distribucionales de las 15 variables (13 atributos y 2 medidas de funcionamiento), particularmente normalidad y curtosis. Se debe emplear el procedimiento descrito en el Capítulo 2 de evaluación de estos supuestos, junto con los procedimientos de curtosis y simetría. En este ejemplo no se encuentra ninguna variable que incumpla significativamente la normalidad o tenga una curtosis pronunciada y, por tanto, las 15 variables se consideran apropiadas para su uso.

Cuando se contrastan una serie de relaciones causales, la matriz de covarianzas es el tipo de matriz de entrada preferido. En este ejemplo, las correlaciones se utilizan tanto por razones prácticas como por razones teóricas. Desde una perspectiva práctica, las correlaciones son más fáciles de interpretar y el diagnóstico de los resultados es más directo. Desde una perspectiva teórica, el principal propósito del análisis es examinar las *pautas* de las relaciones entre los constructos endógenos y exógenos. A estos efectos la matriz de correlación es una matriz de entrada aceptable. El modelo probablemente sufra más de errores de especificación, como es la omisión de variables relevantes del modelo. Tanto la práctica como la experiencia, sugieren la existencia de predictores adicionales de satisfacción de los clientes con HATCO. Por tanto, dada esta limitación, el investigador debería deducir conclusiones sólo acerca de las pautas de las relaciones más que sobre la capacidad predictiva de los constructos [22]. A los efectos fijados, el ejemplo emplea la matriz de correlación (véase Tabla 11.11). El tamaño muestral de 136 clientes cae dentro de los límites aceptables para su uso dentro de SEM.

Estimación del modelo

Como en el ejemplo anterior, el modelo de ecuaciones estructurales se estimó con el programa LISREL [67]. Como se anotó en nuestra discusión del análisis confirmatorio, cada constructo debe hacerse invariante a través de uno de los dos métodos. En este análisis empleamos el método no utilizado anteriormente (es decir, fijamos la ponderación de una variable por constructo con un valor de 1,0) para ilustrar este procedimiento. Los coeficientes de los modelos de estructura y de medida no se ven afectados por el modelo seleccionado. Por tanto, las variables con ponderaciones fijas de 1,0 son X_1, X_7 y X_{10}.

Quinto paso: Evaluación de la identificación del modelo estructural

Antes de examinar los resultados, el investigador debe asegurarse en primer lugar de que se identifica el modelo. El programa LISREL evalúa la identificación del modelo y señala casi todos los problemas. Para este modelo, no se ha indicado ningún error. Como ejemplo de lo que puede provocar un problema de identificación, la especificación de una relación recíproca (doble dirección) entre nivel de uso y satisfacción provocaría en la mayor parte de los casos un problema de identificación con, al menos, una de las ecuaciones estructurales. Aunque las relaciones recíprocas puedan ser estimadas, exigen restricciones al modelo en otros aspectos de las ecuaciones estructurales. Los lectores interesados pueden ir a Bollen [22] o Hayduk [54].

Sexto paso: Evaluación de los criterios de calidad del ajuste

Una vez que se ha asegurado que el modelo está correctamente especificado y el proceso de estimación no está restringido por problemas de identificación, el investigador puede proceder a eva-

TABLA 11.11. Matriz de correlación para el SEM

Variables	Y_1	Y_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
Y_1 Nivel de uso	1,000														
Y_2 Satisfacción	0,411	1,000													
X_1 Calidad del producto	0,288	0,168	1,000												
X_2 Precisión de facturación	0,359	0,159	0,785	1,000											
X_3 Apoyo técnico	0,268	0,141	0,676	0,637	1,000										
X_4 Introducción de nuevos productos	0,212	0,081	0,581	0,622	0,627	1,000									
X_5 Fiabilidad de entrega	0,250	0,060	0,632	0,644	0,538	0,699	1,000								
X_6 Servicio al cliente	0,305	0,127	0,690	0,667	0,551	0,625	0,692	1,000							
X_7 Valor del producto	0,328	0,133	0,293	0,263	0,336	0,290	0,207	0,174	1,000						
X_8 Proveedor precios bajos	0,268	0,046	0,184	0,124	0,230	0,260	0,110	0,254	0,301	1,000					
X_9 Posición negociadora	0,142	0,104	0,289	0,249	0,299	0,219	0,155	0,299	0,307	0,676	1,000				
X_{10} Confluencia de intereses	0,329	0,063	0,327	0,292	0,112	0,292	0,326	0,253	0,179	0,114	0,173	1,000			
X_{11} Integridad y honestidad	0,519	0,077	0,413	0,385	0,265	0,346	0,259	0,262	0,347	0,225	0,205	0,411	1,000		
X_{12} Flexibilidad	0,510	0,090	0,330	0,272	0,192	0,151	0,210	0,225	0,348	0,259	0,162	0,555	0,532	1,000	
X_{13} Resolución de problemas	0,341	0,173	0,154	0,173	0,182	0,153	0,107	0,102	0,272	0,052	0,093	0,202	0,411	0,425	1,000

TABLA 11.12. Resultados SEM: Estimaciones de parámetros estandarizados para el modelo estructural

COEFICIENTES DE ECUACIONES ESTRUCTURALES (VALORES DE *t* ENTRE PARÉNTESIS)

		Constructos endógenos		*Constructos exógenos*			
Constructos endógenos		*Nivel de uso*	*Satisfacción*	*Factores empresa y producto*	*Factores basados en precios*	*Factores de relaciones de compra*	*(R^2) Ajuste de la ecuación estructural*
Nivel de uso	=	0,000	0,000	0,056 (0,621)	0,038 (0,428)	0,615 (4,987)	0,433
Satisfacción	=	0,411 (5,241)	0,000	0,000	0,000	0,000	0,169

CORRELACIÓN ENTRE LOS CONSTRUCTOS EXÓGENOS (VALORES DE *t* ENTRE PARÉNTESIS)

	Constructos exógenos		
Constructos exógenos	*Factores empresa y producto*	*Factores basados en precios*	*Factores de relaciones de compra*
Factores empresa y producto	1,000		
Factores basados en precios	0,355 (2,724)	1,000	
Factores relaciones de compra	0,465 (3,746)	0,353 (2,489)	1,000

luar los resultados específicos del modelo propuesto. Si se cumplen los supuestos subyacentes a SEM, se evalúan los coeficientes estimados junto con el ajuste conjunto del modelo.

Estimaciones infractoras

Una vez que se cumplen los supuestos de SEM, se examinan en primer lugar los resultados en búsqueda de estimaciones sin sentido o teóricamente inconsistentes. Las tres estimaciones infractoras más habituales son las varianzas de error negativas, coeficientes estandarizados que exceden o están muy cerca de 1,0 o errores estándar muy elevados. El examen de los residuos estandarizados en las Tablas 11.12 y 11.13 no revela casos que tenga problemas de este tipo.

Ajuste del modelo conjunto

Antes de evaluar los modelos de medida o estructurales, el investigador debe analizar el ajuste global del modelo para comprobar que es una representación adecuada del conjunto completo de relaciones causales. Se utilizan cada uno de los tres tipos de medidas de calidad del ajuste.

Medidas de ajuste absoluto Las tres medidas más básicas del ajuste absoluto son el ratio de verosimilitud de la chi-cuadrado (χ^2), el índice de bondad del ajuste (GFI) y la raiz cuadrada de la media de los residuos al cuadrado (RMSR) (véase la Tabla 11.14). El valor de la chi-cuadrado de 178,714 con 85 grados de libertad es estadísticamente significativo para el nivel de significación 0,000. Dado que la sensibilidad de esta medida no está demasiado afectada por el tamaño muestral de 136, el investigador puede concluir que existen diferencias significativas. Sin embargo, debemos también anotar que el contraste de la chi-cuadrado se hace más sensible a medida que el número de indicadores aumenta. Con esto en mente, examinaremos otras medidas. El valor GFI de 0,865 es un nivel marginal aceptable, como lo es el valor de RMSR de 0,076. Esta última medida debe

TABLA 11.13 Resultados SEM: estimaciones de parámetros estandarizados para el modelo de medida

PONDERACIONES DE LOS CONSTRUCTOS (VALORES DE LA *t* ENTRE PARÉNTESIS)

	Constructos exógenos		
Indicadores	*Factores empresa y producto*	*Factores basados en precios*	*Factores de relaciones de compra*
X_1 Calidad del producto	0,863 (0,000)[a]	0,000	0,000
X_2 Precisión de facturación	0,857 (12,855)	0,000	0,000
X_3 Apoyo técnico	0,747 (10,297)	0,000	0,000
X_4 Introducción de nuevos productos	0,759 (10,550)	0,000	0,000
X_5 Fiabilidad de entrega	0,781 (11,020)	0,000	0,000
X_6 Servicio al cliente	0,803 (11,524)	0,000	0,000
X_7 Valor del producto	0,000	0,404 (0,000)[a]	0,000
X_8 Proveedor precios bajos	0,000	0,811 (4,181)	0,000
X_9 Posición negociadora	0,000	0,821 (4,169)	0,000
X_{10} Confluencia de intereses	0,000	0,000	0,603 (0,000)[a]
X_{11} Integridad y honestidad	0,000	0,000	0,729 (6,216)
X_{12} Flexibilidad	0,000	0,000	0,786 (6,447)
X_{13} Resolución de problemas	0,000	0,000	0,506 (4,786)

[a] Valores no calculados porque la ponderación se fijó en 1,0 para fijar la varianza del constructo.

evaluarse a la luz de la matriz de correlación de entrada, que indica que en este contexto es un valor relativamente alto.

Como complemento de estas tres medidas básicas, el investigador también puede examinar otras medidas de ajuste absoluto. El error de aproximación cuadrático medio (KMSA) tiene un valor de 0,090, que cae justo fuera del rango aceptable de 0,08 o menor, pero es más pequeño que el umbral superior de 0,10. Se pueden utilizar otras tres medidas —el índice de no centralidad (NCP), el índice de no centralidad a escala y el índice de validación cruzada esperada (ECVI)— en la comparación entre modelos alternativos. No hay un rango de valores aceptados para estas medidas. Se examinan después, cuando se analizan los modelos rivales.

Todas las medidas de ajuste absoluto indican que el modelo es como mucho marginalmente aceptable. Sin embargo, esto no debería impedir un examen mas detenido de los resultados, a no ser que el ajuste conjunto del modelo sea tan malo que haya que considerar no aptos tanto los modelos de medida como los estructurales. Además, los otros tipos de medidas del ajuste ofrecerán diferentes perspectivas sobre la aceptabilidad del ajuste del modelo.

Medidas de ajuste incremental Además de las medidas conjuntas del ajuste, se puede evaluar el modelo en comparación con un modelo nulo. En este caso, el modelo nulo es un modelo de factor único sin error de medida. El modelo nulo tiene un valor de chi-cuadrado de 1.040,149 con 105 grados de libertad. Aunque ganemos en una sustancial reducción en el valor de

TABLA 11.14. Medidas de bondad del ajuste para SEM

Medidas ofrecidas por LISREL	
Medidas de ajuste absoluto	
Chi-cuadrado del modelo estimado (χ^2)	178,714[a]
Grados de libertad	85
Nivel de significación	0,000[a]
Parámetro de no centralidad (NCP)	93,714
Índice de bondad del ajuste (GFI)	0,865[a]
Residuo cuadrático medio (RMSR)	0,076[a]
Error de aproximación cuadrático medio (RMSEA)	0,090
Valor-*P* de ajuste cercano (RMSEA < 0,05)	0,000
Índice de validación cruzada esperada (ECVI)	
ECVI para modelo estimado	1,842
ECVI para modelo saturado	1,778
ECVI para modelo independiente	7,927
Medidas de ajuste incremental	
Chi-cuadrado (χ^2) de modelo nulo o independiente	1.040,194
Grados de libertad	105
Índice de bondad del ajuste ajustado (AGFI)	0,810[a]
Índice Tucker-Lewis (TLI) o índice de ajuste no-normado (NNFI)	0,876
Índice de ajuste normado (NFI)	0,828
Medidas de ajuste de parsimonia	
Índice de ajuste normado de parsimonia (PNFI)	0,670
Índice de bondad del ajuste de parsimonia	0,613
Criterio de información de Akaike (AIC)	
Modelo estimado	248,714
Modelo saturado	240,000
Modelo independiente	1.070,194
Índice de ajuste comparativo (CFI)	0,900
Índice de ajuste incremental (IFI)	0,902
Índice de ajuste relativo (RFI)	0,788
N crítica (CN)	90,318
Tamaño de la muestra: 136 encuestados	
Medidas calculadas de bondad del ajuste del modelo conjunto	
Chi-cuadrado normada	$\frac{178,714}{85} = 2,103$
Parámetro de no-normalidad medida (SNCP)	$\frac{93,174}{136} = 0,689$

[a] Medidas ofrecidas directamente por todas las versiones de LISREL. Las otras medidas sólo se ofrecen en la versión 8 de LISREL y deben ser calculadas por separado en las versiones anteriores.

la chi-cuadrado debido a los coeficientes estimados del modelo, las medidas de ajuste incremental también ofrecen evidencia adicional. El AGFI, el índice Tucker-Lewis y el NFI caen ligeramente por debajo del umbral deseado de 0,90.

Aunque el umbral de 0,90 no tiene fundamento estadístico, la experiencia y la investigación práctica han demostrado su utilidad en la distinción entre los modelos aceptables y no aceptables. Sin embargo, todas las medidas del ajuste incremental exceden de 0,80, y el verdadero contraste proviene de la comparación del modelo propuesto frente a los modelos alternativos o rivales.

Medidas de ajuste de parsimonia Este último tipo de medidas ofrece un punto de partida para la comparación entre modelos de diferente complejidad y objetivos. Una medida aplicable para la evaluación de un único modelo es la medida de la chi-cuadrado normada. Con un valor calculado de 2,103, cae dentro de los umbrales límites de esta medida pero excede otros límites. Por tanto, de nuevo sólo se aporta evidencia marginal. Existen otras tres medidas de ajuste de

parsimonia (el índice de ajuste normado de parsimonia, el índice de bondad del ajuste de parsimonia y el criterio de información de Akaike), pero sólo son apropiadas para las comparaciones entre modelos.

Una revisión de los tres tipos de medidas de ajuste conjunto revela una pauta consistente de evidencia marginal del modelo conjunto tal y como se ha propuesto. Como se ha apuntado previamente, el verdadero contraste del modelo conjunto es una comparación de una serie de modelos alternativos, que se lleva a cabo en el séptimo paso.

Ajuste del modelo de medida

Incluso aunque se encontrara evidencia marginal para el modelo conjunto, se consideró suficiente proceder a evaluar también el ajuste del modelo de medida. El primer paso es un examen de las ponderaciones, centrándose particularmente en cualquier ponderación no significativa. En referencia a la Tabla 11.13, vemos que todos los indicadores son estadísticamente significativos para los constructos propuestos. Dado que ningún indicador tiene ponderaciones tan bajas como para ser destruidas y el modelo reestimado, se necesita calcular la fiabilidad y las medidas de varianza extraída.

La Tabla 11.15 contiene los cálculos, tanto de la fiabilidad como de las medidas de varianza extraída. En términos de fiabilidad, los tres constructos exógenos exceden el nivel sugerido de 0,70. En términos de varianza extraída, el constructo empresa / producto excede el valor umbral de 0,50. El constructo basado en el precio sólo se queda un poco por debajo de 0,50 y el constructo de la relación-compra tiene algún valor más bajo de (0,442). Por tanto, para los tres constructos, los indicadores son suficientes en términos de cómo se especifica el modelo de medida.

Ajuste del modelo estructural

Una vez evaluado el modelo conjunto y los aspectos del modelo de medida, el investigador se encuentra ahora preparado para examinar los coeficientes estimados, tanto para implicaciones prácticas como teóricas. La revisión de la Tabla 11.12 revela que ambas ecuaciones estructurales contienen coeficientes estadísticamente significativos. Dado que las relaciones causales vinculan las tres dimensiones de evaluación con el nivel de uso, sólo una dimensión, la relación de compra, es estadísticamente significativa. Por tanto, los aspectos más personales de la transacción, tales como reciprocidad, integridad, flexibilidad en los contratos y resolución de problemas, tienen un único y sustantivo impacto sobre los aumentos del nivel de uso de producto. Por tanto, aunque HATCO no debe ignorar los otros aspectos del negocio, se debe hacer más énfasis en el mantenimiento de las relaciones de compra y de desarrollo de las nuevas centrándose en estas características. Además, el efecto combinado de estos tres valores consigue un valor de R^2 de 0,433, o 43,3 por ciento de la varianza de los niveles de uso. Dado que la ecuación estructural predice la satisfacción, el otro constructo exógeno (nivel de uso) es estadísticamente significativo. Por tanto, se ha identificado una relación causal significativa y puede utilizarse ahora como base para la formulación de modelos alternativos o rivales. Se espera un bajo valor del R^2 (0,169) es esperado porque se basa sólo en la correlación entre el nivel de uso y satisfacción.

También son de interés las correlaciones entre las dimensiones de evaluación (constructos exógenos). En este ejemplo, cada constructo está estadísticamente correlacionado con los otros constructos exógenos. Esta correlación revela que, mientras que la relación de compra es fundamental para incrementar los niveles de uso, los tres constructos se van entretejiendo, y la empresa no debe centrarse exclusivamente en una única dirección.

Como último medio de examinar los resultados, se deberían revisar las correlaciones entre los constructos estimados (variables endógenas) y los valores elevados deberían ser tomados como una indicación de un nivel inaceptable de intercorrelación de los constructos. Para los constructos endógenos de este estudio, la correlación de sus valores estimados es de sólo 0,411 (véase Tabla 11.12), muy baja como para indicar una intercorrelación problemática.

Modelos rivales

El enfoque final para la evaluación del modelo es comparar el modelo propuesto con una serie de modelos rivales, que actúan como explicaciones alternativas del modelo propuesto. De esta forma, el investigador puede determinar si el modelo propuesto es aceptable, a pesar del ajuste conjunto (dentro de los límites razonables) porque ningún otro modelo formulado de forma parecida

TABLA 11.15. Fiabilidad y estimaciones de varianza extraída para los constructos exógenos en SEM

Fiabilidad

$$\text{Fiabilidad de constructo} = \frac{(\text{Suma de las ponderaciones estandarizadas})^2}{(\text{Suma de las ponderaciones estandarizadas})^2 + (\text{Suma del error de medida del indicador})}$$

Suma de las ponderaciones estandarizadas

Factores producto y empresa = 0,863 + 0,857 + 0,747 + 0,759 + 0,781 + 0,803 = 4,81
Factores basados en precio = 0,404 + 0,811 + 0,821 = 2,036
Relaciones de compra = 0,603 + 0,729 + 0,786 + 0,506 = 2,624

Suma del error de medida[a]

Factores de empresa y producto = 0,256 + 0,265 + 0,442 + 0,424 + 0,390 + 0,355 = 2,132
Factores basados en precio = 0,837 + 0,342 + 0,327 = 1,506
Relaciones de compra = 0,636 + 0,469 + 0,382 + 0,744 = 2,231

Cálculo de la fiabilidad

$$\text{Factores de empresa y producto} = \frac{(4{,}81)^2}{(4{,}81)^2 + 2{,}132} = 0{,}916$$

$$\text{Factores basados en precio} = \frac{(2{,}036)^2}{(2{,}036)^2 + 1{,}506} = 0{,}734$$

$$\text{Relaciones de compra} = \frac{(2{,}624)^2}{(2{,}624)^2 + 2{,}231} = 0{,}755$$

Varianza extraída

$$\text{Varianza extraída} = \frac{\text{Suma de los cuadrados de las ponderaciones estandarizadas}}{\text{Suma de los cuadrados de las ponderaciones estandarizadas} + \text{Suma de error de medida del indicado}}$$

Suma de los cuadrados de las ponderaciones estandarizadas

Factores de empresa y producto = $0{,}863^2 + 0{,}857^2 + 0{,}747^2 + 0{,}759^2 + 0{,}781^2 + 0{,}803^2 = 3{,}868$
Factores basados en precio = $0{,}404^2 + 0{,}811^2 + 0{,}821^2 = 1{,}495$
Relaciones de compra = $0{,}603^2 + 0{,}729^2 + 0{,}786^2 + 0{,}502^2 = 1{,}769$

Cálculo de la varianza extraída

$$\text{Factores de empresa y producto} = \frac{3{,}868}{3{,}868 + 2{,}132} = 0{,}645$$

$$\text{Factores basados en precio} = \frac{1{,}495}{1{,}495 + 1{,}506} = 0{,}498$$

$$\text{Relaciones de compra} = \frac{1{,}769}{1{,}769 + 2{,}231} = 0{,}442$$

[a] El error de medida del indicador puede ser calculado como 1 – (ponderación estandarizada)^2 o la diagonal de la matriz de correlación del error de media (matriz delta-teta) en resultado LISREL.

puede conseguir un elevado nivel de ajuste. Este paso es particularmente importante cuando el estadístico de la chi-cuadrado indica que no existen diferencias significativas en el ajuste conjunto del modelo, porque siempre puede haber un modelo mejor ajustado, incluso en el caso de diferencias no significativas. Para un enfoque sistemático de especificación de modelos rivales, véase Anderson y Gerbing [4].

A efectos de este ejemplo, se proponen dos modelos alternativos (véase Figura 11.10 para ver estos modelos expresados como diagramas de relaciones). El primer modelo (COMPMOD1) añade tres constructos exógenos como predictores de satisfacción. El segundo modelo (COMPMOD2) sólo añade una relación significativa de niveles de uso. Como forma de comparación, se calculan un conjunto de medidas de calidad del ajuste para cada modelo y a continuación se comparan para determinar cuál de las tres es la más parsimoniosa.

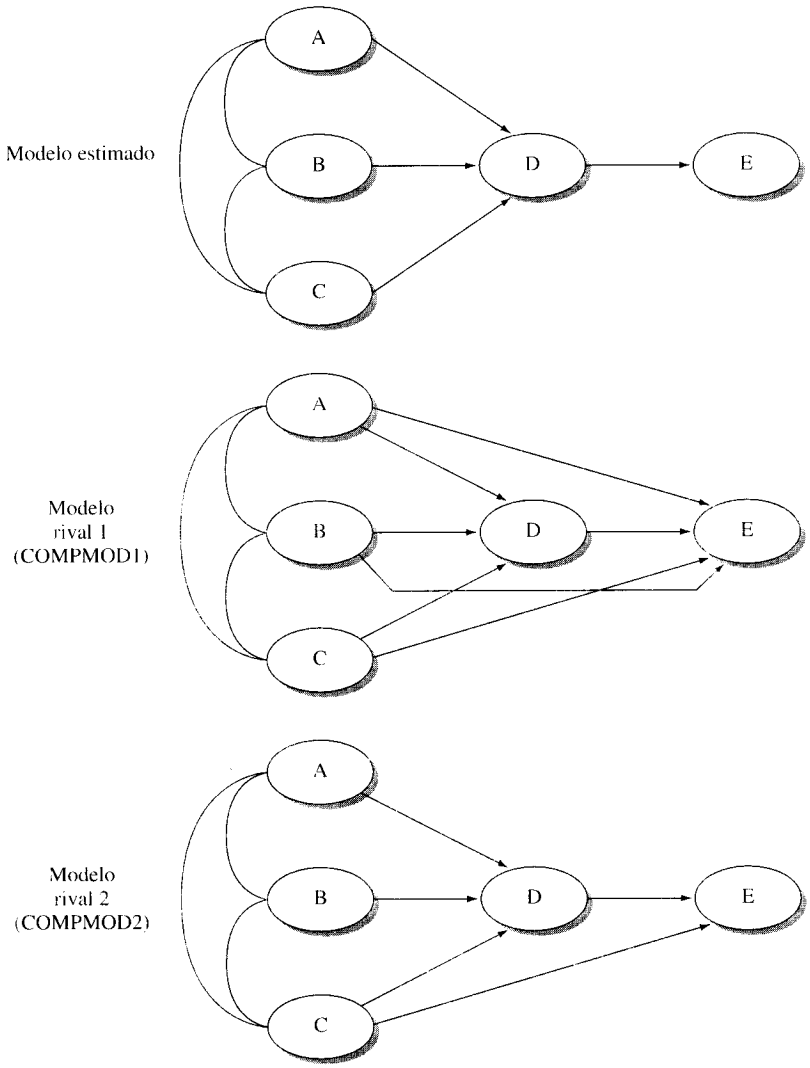


FIGURA 11.10. Diagramas de secuencia del modelo estimado y los modelos rivales (COMPMOD1 y COMPMOD2).

La Tabla 11.16 compara los tres modelos con los tres tipos de medidas. Para las medidas de ajuste absoluto, COMPMOD1 tiene el valor más bajo de la chi-cuadrado, pero este modelo también tiene el mayor número de parámetros estimados (y por tanto los menores grados de libertad). COMPMOD1 es también el más bajo con la medida RMSR. COMPMOD2 sobresale en las medidas que intentan minimizar el tamaño muestral y la estimaciones específicas de la muestra del ajuste del modelo. La Tabla 11.16 muestra que los índices NCP, SNCP y ECVI tienen todas sus valores más bajos con COMPMOD2. El modelo estimado, aunque no consigue el mejor ajuste de cualquiera de estas medidas, está todavía muy cerca de todas ellas, siendo una alternativa viable para su aceptación junto con los otros modelos rivales.

Dos de las medidas de ajuste incremental favorecen COMPMOD2 (AGFI y NNFI), mientras que COMPMOD1 tiene el mejor comportamiento de la medida NFI. De nuevo, el modelo estimado está bastante cerca, sin diferencias sustantivas. Las medidas de ajuste de parsimonia son el conjunto final a considerar. Aquí, el modelo estimado tiene el mejor ajuste tal y como se mide por el PNFI y PGFI, y COMPMOD2 sobresale en la chi-cuadrado normada y en el criterio de información de Akaike.

Los resultados para los tres tipos de medidas muestran resultados mixtos, a veces favoreciendo el modelo estimado o uno de los modelos rivales. Si el interés se limita a los índices de parsimonia, que tiene en cuenta la parsimonia del modelo, los resultados todavía están escindidos entre el modelo original estimado y el segundo modelo rival (COMPMOD2). Por tanto, aunque el ajuste del modelo propuesto no excede las pautas recomendadas en muchos casos, lo aceptamos con reservas hasta que se puedan añadir constructos adicionales, medidas refinadas o relaciones causales reespecificadas. La selección de uno de los modelos rivales debe hacerse tanto sobre bases teóricas como empíricas, al igual que se hizo en la formulación del modelo teórico original que era la base del modelo estimado.

Este ejemplo ilustra el valor la contrastación de los modelos rivales para todos los modelos de ecuaciones estructurales, con el fin de que el investigador se asegure de que el modelo propuesto revisado es verdaderamente el «mejor» modelo disponible.

TABLA 11.16. Comparación de medidas de bondad del ajuste para el modelo estimado y para los modelos rivales

Medidas de bondad del ajuste	*Modelo estimado*	*Modelos rivales*	
		$COMPMOD_1$	$COMPMOD_2$
Medidas de ajuste absoluto			
Ratio de verosimilitud de la chi-cuadrado (χ^2)	178,714	174,450	175,397
Grados de libertad	85	82	84
Parámetro de no centralidad (NCP)	93,714	92,450	91,397
Parámetro de no centralidad a medida (SNCP)	0,689	0,680	0,672
Índice de bondad del ajuste (GFI)	0,865	0,867	0,866
Residuo cuadrático medio (RMSR)	0,076	0,074	0,075
Error de aproximación cuadrático medio (RMSEA)	0,090	0,091	0,090
Índice de validación cruzada esperado (ECVI)	1,842	1,855	1,833
Medidas del ajuste incremental			
Índice de bondad del ajuste incremental (AGFI)	0,810	0,805	0,809
Índice Tucker-Lewis (TLI) o (NNFI)	0,876	0,873	0,878
Índice de ajuste incremental (NFI)	0,828	0,832	0,831
Medidas de ajuste de parsinomia			
Índice de ajuste normado de parsinomia (PNFI)	0,670	0,650	0,665
Índice de bondad del ajuste de parsinomia (PGFI)	0,613	0,592	0,606
Chi-cuadrado normada	2,103	2,127	2,088
Criterio de información de Akaike (AIC)	248,714	250,450	247,397

Séptimo paso: Interpretación y modificación del modelo

El último paso consiste en interpretar los resultados en términos empíricos y prácticos, así como en examinar de los resultados de cualquier modificación potencial del modelo. Aunque la interpretación es el fundamento para apoyar el modelo teórico, se necesitan procedimientos de reespecificación del modelo para comprobar que el modelo estimado tiene las restricciones empíricas más apropiadas para maximizar la calidad del ajuste del modelo.

Interpretación

El modelo de medida ofrece al investigador un medio de dar significado a los constructos, en este ejemplo los tres constructos exógenos, dado que los dos constructos endógenos eran indicadores aislados de los constructos. Cada uno de los indicadores de los tres conjuntos tiene indicadores claramente comparables, con la excepción hecha de la ponderación de X_7 sobre la base del constructo basado en el precio (véase Tabla 11.13). Las ponderaciones son todas estadísticamente significativas, aportando por tanto el fundamento teórico para la evaluación de los indicadores para cada constructo.

La significación de los test de los parámetros del modelo estructural es la base para aceptar o rechazar las relaciones propuestas entre los constructos endógenos y exógenos. En la primera relación básica, se propusieron como antecedentes del uso los tres constructos exógenos (factores empresa / producto, factores basados en los precios y factores de relaciones de compra). El modelo estimado, sin embargo, sólo encontró evidencia para el papel de las relaciones de compra al afectar al nivel de uso. Los otros constructos tenían estimaciones de los parámetros muy bajas y éstos fueron considerados también significativos y todos ellos positivos. La segunda relación hacía referencia a la predicción de la satisfacción, que fue considerada como una función únicamente del nivel de uso. Esta relación estaba apoyada por un parámetro de 0,411 (véase Tabla 11.12), que era estadísticamente significativa y de la dirección apropiada (positiva). La discusión anterior de los modelos rivales identificaba ciertos efectos adicionales que podrían afectar a la satisfacción, pero que debían ser contrastados en ulteriores investigaciones.

Reespecificación del modelo

Finalmente, examinamos algunos elementos de diagnóstico (índices de residuo y modificación) que pueden indicar modificaciones del modelo potencialmente significativas. Como ya se anotó, estamos suponiendo que todas las modificaciones del modelo tienen apoyo teórico antes de ser instrumentadas.

Análisis de los residuos normalizados La Tabla 11.17 contiene los residuos normalizados del modelo propuesto cuando se estima mediante LISREL. Existen 11 residuos potencialmente significativos –residuos que exceden el valor umbral del ± 2,58, que excede de la pauta de tener sólo un 5 por ciento de residuos normalizados que superan el valor umbral, que en este caso serían seis residuos (5 por ciento de 120). Podemos anotar sólo una pauta consistente entre los residuos. La variable X_7 está conectada con la mayoría de los residuos que exceden de 2,58. Esta observación indica que quizá la eliminación de sólo este único indicador creará una mejora sustancial del ajuste del modelo. Si se realiza esta modificación, el valor de la chi-cuadrado disminuye hasta 146,94 con 71 grados de libertad. Utilizando un índice de ajuste de parsimonia, notamos que el valor calculado para este modelo de indicador reducido es 0,557, que no es mayor que el modelo propuesto del original. Por tanto, aunque el ajuste del modelo mejorará, no lo hará la parsimonia del modelo. Si excluimos X_7, los residuos normalizados cumplirán marginalmente la pauta. La decisión del investigador en relación con la exclusión de X_7 debería basarse en fundamentos teóricos.

TABLA 11.17. Residuos normalizados de SEM

	Y_1	Y_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
Y_1 Nivel de uso	0,000														
Y_2 Satisfacción	0,000	0,000													
X_1 Calidad del producto	–0,542	0,546	0,000												
X_2 Precisión de facturación	1,526	0,436	3,138	0,000											
X_3 Apoyo técnico	0,052	0,409	1,430	–0,154	0,000										
X_4 Introducción de nuevos productos	–1,196	–0,384	–3,423	–1,292	1,839	0,000									
X_5 Fiabilidad de entrega	–0,600	–0,693	–2,039	–1,233	–1,495	3,569	0,000								
X_6 Servicio al cliente	0,437	0,124	–0,120	–1,117	–1,724	0,559	2,499	0,000							
X_7 Valor del producto	2,972	1,040	2,293	1,887	2,985	2,363	1,243	0,780	0,000						
X_8 Proveedor precios bajos	1,402	–0,593	–1,454	–2,736	0,255	0,757	–2,176	0,438	–1,103	0,000					
X_9 Posición negociadora	–2,788	0,147	0,870	–0,013	1,469	–0,034	–1,383	1,287	–1,058	3,810	0,000				
X_{10} Confluencia de intereses	–1,582	–1,347	1,396	0,851	–1,468	1,199	1,641	0,427	1,174	–0,941	–0,035	0,000			
X_{11} Integridad y honestidad	1,385	–1,704	2,357	1,830	0,188	1,492	–0,096	–0,181	3,186	0,308	–0,128	–0,847	0,000		
X_{12} Flexibilidad	–0,189	–1,794	0,310	–0,900	–1,450	–2,287	–1,146	–1,323	3,150	0,735	–1,478	2,956	–2,310	0,000	
X_{13} Resolución de problemas	0,214	0,485	–0,742	–0,436	0,085	–0,370	–1,105	–1,271	2,467	–1,343	–0,775	–1,977	1,063	0,836	0,000

Índices de modificación Otra indicación de la posible reespecificación del modelo es el índice de modificación. La Tabla 11.18 contiene los índices de modificación obtenidos durante la estimación del modelo propuesto. Buscando los valores que exceden 3,84 se puede ver que las modificaciones se sugieren sólo en el modelo de medida. Deberíamos tener en cuenta que los dos índices de modificación más altos se dan para X_7, la variable que el análisis residual sugirió eliminar. Además de X_7, se sugirieron ponderaciones sobre más de un constructo diferenciado para X_8 y X_9, aunque la reducción anticipada en la chi-cuadrado es mínima y no es apropiada para la reespecificación del modelo.

TABLA 11.18 Índices de modificación para SEM

MODELO DE MEDIDA

Indicadores de constructos exógenos

	Constructos exógenos		
Variables	*Factores empresa y producto*	*Factores basados en precios*	*Factores de relaciones de compra*
X_1 Calidad del producto	0,000	0,040	2,224
X_2 Precisión de facturación	0,000	1,574	0,690
X_3 Apoyo técnico	0,000	2,009	0,872
X_4 Introducción de nuevos productos	0,000	0,543	0,276
X_5 Fiabilidad de entrega	0,000	3,565	0,570
X_6 Servicio al cliente	0,000	1,294	0,589
X_7 Valor del producto	6,242	0,000	15,289
X_8 Proveedor precios bajos	3,907	0,000	0,000
X_9 Posición negociadora	0,489	0,000	3,916
X_{10} Confluencia de intereses	1,094	0,107	0,000
X_{11} Integridad y honestidad	2,959	0,574	0,000
X_{12} Flexibilidad	3,484	0,003	0,000
X_{13} Resolución de problemas	0,778	0,766	0,000

MODELO ESTRUCTURAL

	Constructos endógenos		*Constructos exógenos*		
Constructos endógenos	*Nivel de uso*	*Satisfacción*	*Factores empresa y producto*	*Factores basados en precios*	*Factores de relaciones de compra*
Nivel de uso	0,000	2,729	0,000	0,000	0,000
Satisfacción	0,000	0,000	0,016	0,026	3,249

CORRELACIÓN ENTRE LOS CONSTRUCTOS EXÓGENOS

	Constructos exógenos		
	Factores de empresa y producto	*Factores basados en precios*	*Factores de relaciones de compra*
Factores de empresa y producto	0,000		
Factores basados en precios	0,000	0,000	
Factores relaciones de compra	0,000	0,000	0,000

Correlación entre los constructos endógenos (Ecuaciones estructurales)

	Nivel de uso	*Satisfacción*
Nivel de uso	0,000	
Satisfacción	2,729	0,000

Repaso del proceso de siete pasos

El proceso de siete pasos ha servido para investigar empíricamente una serie de relaciones casuales con constructos (endógenos) dependientes interrelacionados. El modelo estimado, incluso aunque no consiguiera los niveles recomendados de ajuste, puede representar el mejor modelo disponible hasta que una ulterior investigación determine las mejoras en las relaciones teóricas o medidas de constructos. Aunque diversas modificaciones del modelo mejoran su ajuste, no aumentan la parsimonia del modelo, y se puede encontrar poco apoyo teórico para estas reespecificaciones de la teoría. Por tanto, en situaciones como estas, el investigador debería ser cauteloso sobre el tipo y la extensión de las reespecificaciones del modelo consideradas.

Resumen

La modelización de ecuaciones estructurales es una técnica que combina elementos tanto de la regresión múltiple como del análisis factorial. Permite al investigador no sólo evaluar las muy complejas interrelaciones de dependencia sino también incorporar los efectos del error de medida sobre los coeficientes estructurales al mismo tiempo. Hasta la finalización de este capítulo, tanto el investigador académico como el analista empresarial deberían ser capaces de reconocer los beneficios aportados por SEM. Aunque SEM es útil en muchos casos, debería ser utilizado sólo de modo confirmatorio, dejando los análisis exploratorios a otras técnicas multivariantes.

Cuestiones de revisión

1. ¿Cuáles son las similitudes entre el modelo de ecuaciones estructurales y las técnicas multivariantes discutidas en los capítulos anteriores?
2. ¿Por qué debería el investigador evaluar el error de medida e incorporarlo al análisis?
3. Utilizando la definición de «teoría» presentada en este capítulo, describa las teorías que puedan ser de interés tanto para los investigadores académicos como para los empresariales.
4. Especificar un conjunto de relaciones causales y a continuación representarlas en un diagrama de relaciones.
5. ¿Cuáles son los criterios por los cuales el investigador debería decidir si un modelo tiene un nivel de ajuste aceptable?
6. Utilizando el diagrama de relaciones para la cuestión 4 (o una preparada para esta cuestión), especificar al menos dos modelos alternativos que aporten lógica teórica al primer modelo especificado.
7. ¿Qué quiere decir que el análisis factorial confirmatorio tiene un orden elevado (segundo o tercer orden)? ¿Cuándo debería utilizarse y qué significa que otros análisis factoriales no lo hagan?

APÉNDICE 11A

Una representación matemática en notación LISREL

Hasta este momento la discusión de la modelización de las ecuaciones estructurales ha descansado en una terminología similar a la utilizada en nuestra discusión de la regresión múltiple y de otras relaciones de dependencia. Sin embargo, un investigador que intente (1) leer manuales de programas, artículos o textos sobre este tema, o (2) publicar en revistas académicas utilizando esta técnica, necesitará una introducción más formal que la notación utilizada en SEM. Dada su amplia aplicación, LISREL [64, 65, 67] se ha convertido en la notación estándar. Este apéndice presenta en primer lugar la formulación completa del conjunto de relaciones causales y relaciones de medida en notación LISREL. A continuación, con un ejemplo se irá del diagrama de secuencias a las ecuaciones estructurales en la apropiada notación LISREL de especificación del modelo.

Notación LISREL

El modelo LISREL completo se puede expresar en términos de ocho matrices, dos definiendo las ecuaciones estructurales, dos definiendo la correspondencia de indicadores y constructos, una para la correlación de constructos exógenos, una para la correlación de los constructos endógenos y finalmente dos detallando los errores correlacionados para la medición de las variables endógenas y exógenas. La Tabla 11 A.1 enumera cada una de las matrices así como los constructos e indicadores, y ofrece una breve descripción y notación de la matriz y sus elementos.

Las matrices se utilizan para formar las ecuaciones básicas tanto para los modelos de medida como para los estructurales. La ecuación del modelo estructural es

$$\eta = \Gamma\xi + B\eta + \zeta$$

TABLA 11 A.1. Matrices, constructos e indicadores y notación de la ecuación del modelo LISREL

Modelo LISREL *Elemento*	*Descripción*	*Notación* *Matriz*	*Elemento*
Matriz			
Modelo estructural			
Beta	Relaciones de constructos endógenos a exógenos	B	β_{nn}
Gamma	Relaciones de constructos exógenos a endógenos	Γ	γ_{nm}
Phi	Correlaciones entre constructos exógenos	Φ	ϕ_{mm}
Psi	Correlación de ecuaciones estructurales o constructos endógenos	Ψ	ψ_{nn}
Modelo de medida			
Lambda-X	Correspondencia (ponderaciones) de indicadores exógenos	Λ_x	λ^x_{pm}
Lambda-Y	Correspondencia (ponderaciones) de indicadores exógenos	Λ_y	λ^y_{qn}
Theta-delta	Matriz correlacional de error de predicción para indicadores de constructos exógenos	Θ_δ	θ^δ_{pp}
Theta-epsilon	Matriz correlacional de error de predicción para indicadores de constructos endógenos	Θ_ε	θ^ε_{qq}
Constructos e indicadores			
Constructo			
Exógeno	Constructo exógeno		ξ
Endógeno	Constructo endógeno		η
Indicador			
Exógeno	Indicador exógeno		X
Endógeno	Indicador endógeno		Y
Ecuaciones de modelo de medida y estructural			
Modelo estructural	Relaciones entre constructos endógenos y exógenos	$\eta = \Gamma\xi + \beta\eta + \zeta$	
Modelo de medida			
Exógeno	Especificación de indicadores de constructos exógenos	$X = \Lambda_x\xi + \delta$	
Endógeno	Especificación de indicadores de constructos endógenos	$Y = \Lambda_y\eta + \varepsilon$	
Subíndices de dimensión para matriz			
	Número de constructos exógenos	m	
	Número de constructos endógenos	n	
	Número de indicadores de constructo exógenos	p	
	Número de indicadores de constructo endógenos	q	

Y las ecuaciones del modelo de medida son las siguientes:

Constructos exógenos	Constructos endógenos
$X = \Lambda_x\xi + \delta$	$Y = \Lambda_y\eta + \varepsilon$

Una presentación más directa, sin embargo, es a través de las ecuaciones efectivas (estructurales y de medida). A efectos ilustrativos, existen dos constructos endógenos ($n = 2$) y tres constructos exógenos ($m = 3$), con cuatro indicadores de constructos tanto endógenos como endógenos ($p = 4$, $q = 4$). No se expresan relaciones causales; más bien se muestran las ecuaciones completas (véase Tabla 11A.2). En el ejemplo siguiente ilustra una especificación de un modelo de relaciones propuesto en términos de LISREL.

Cuando las ecuaciones estructurales y de medida están estimadas, los términos de error se hacen parte de las matrices entre-constructo y entre-indicador descritas. Por ejemplo, los valores δ_p y ε_p se estiman realmente como los valores diagonales de las matrices theta-delta y theta-epsilon, respectivamente. Lo mismo, los errores de la ecuación estructural (ζ_n) son las diagonales de la matriz psi.

TABLA 11 A.2. Ecuaciones completas del modelo de medida y estructural

ECUACIONES DEL MODELO ESTRUCTURAL

Constructo endógeno		*Constructo exógeno*		*Constructos endógenos*		*Error*
η_1	=	$\gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3$	+	$\beta_{11}\eta_1 + \beta_{12}\eta_3$	+	ζ_1
η_2	=	$\gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3$	+	$\beta_{21}\eta_2 + \beta_{22}\eta_2$	+	ζ_2

ECUACIONES DEL MODELO DE MEDIDA

Indicador exógeno		*Constructos exógenos*		*Error*
X_1	=	$\lambda^x_{11}\xi_1 + \lambda^x_{12}\xi_2 + \lambda^x_{13}\xi_3$	+	δ_1
X_2	=	$\lambda^x_{21}\xi_1 + \lambda^x_{22}\xi_2 + \lambda^x_{23}\xi_3$	+	δ_2
X_3	=	$\lambda^x_{31}\xi_1 + \lambda^x_{32}\xi_2 + \lambda^x_{33}\xi_3$	+	δ_3
X_4	=	$\lambda^x_{41}\xi_1 + \lambda^x_{42}\xi_2 + \lambda^x_{43}\xi_3$	+	δ_4

Indicador endógeno		*Constructos endógenos*		*Error*
Y_1	=	$\lambda^y_{11}\eta_1 + \lambda^y_{12}\eta_2$	+	ε_1
Y_2	=	$\lambda^y_{21}\eta_1 + \lambda^y_{22}\eta_2$	+	ε_2
Y_3	=	$\lambda^y_{31}\eta_1 + \lambda^y_{32}\eta_2$	+	ε_3
Y_4	=	$\lambda^y_{41}\eta_1 + \lambda^y_{42}\eta_2$	+	ε_4

CORRELACIONES DE ECUACIÓN ESTRUCTURAL ENTRE CONSTRUCTOS

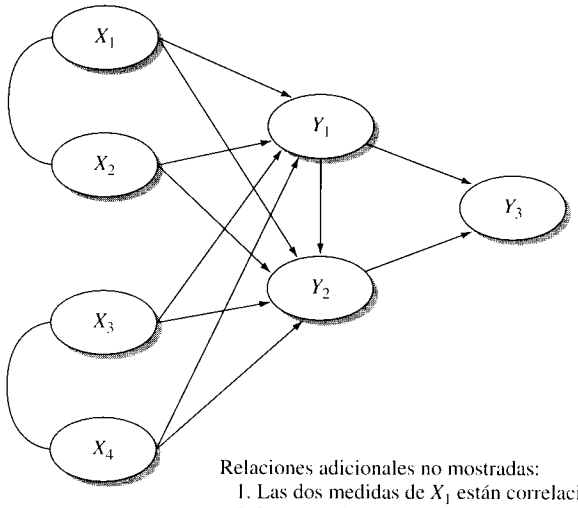
	Entre constructos exógenos (Phi ϕ)			*Entre constructos endógenos (Psi ψ)*	
	ξ_1	ξ_2	ξ_3	η_1	η_2
ξ_1	—				
ξ_2	ϕ_{21}	—		η_1	
ξ_3	ϕ_{31}	ϕ_{32}	—	η_2	ψ_{21} —

CORRELACIONES (INDICADORES) DEL MODELO DE MEDIDA

	Entre indicadores exógenos (Theta-delta θ_δ)				*Entre indicadores endógenos (Theta-epsilon ϕ_ε)*				
	X_1	X_2	X_3	X_4		Y_1	Y_2	Y_3	Y_4
X_1	—				Y_1	—			
X_2	$\phi_{\delta_{21}}$	—			Y_2	$\phi_{\varepsilon_{21}}$	—		
X_3	$\phi_{\delta_{31}}$	$\phi_{\delta_{32}}$	—		Y_3	$\phi_{\varepsilon_{31}}$	$\phi_{\varepsilon_{32}}$	—	
X_4	$\phi_{\delta_{41}}$	$\phi_{\delta_{42}}$	$\phi_{\delta_{43}}$	—	Y_4	$\phi_{\varepsilon_{41}}$	$\phi_{\varepsilon_{42}}$	$\phi_{\varepsilon_{43}}$	—

Desde un diagrama de relaciones a la notación LISREL

Ilustramos ahora el proceso completo de avance desde el diagrama de secuencias hasta la notación completa de LISREL, lista para introducirse en cualquiera de los programas SEM. El diagrama de relaciones de la Figura 11A.1 describe el conjunto de relaciones causales a examinar con este ejemplo. Existen tres variables endógenas (Y o dependientes) en relación con cuatro variables exógenas (X o independientes). Cada uno de los siete constructos se miden por dos variables (V_1 a V_{14}). También, las medidas de X_1 están correlacionadas, como lo están las dos medidas de Y_2. Además, las correlaciones también existen entre dos pares de variables independientes (X_1-X_2 y X_3-X_4) y dos variables dependientes (Y_1-Y_3). No se debe confundir el uso de las variables X e Y en el diagrama de relaciones con la designación de los indicadores endógenos y exógenos, respectivamente, en la notación LISREL. Los indicadores del diagrama de relaciones, que se muestran en la Figura 11A.1, son las variables V_1 a V_{14}.



Relaciones adicionales no mostradas:
1. Las dos medidas de X_1 están correlacionadas
2. Las ecuaciones para Y_1 e Y_3 están correlacionadas
3. Las dos medidas de Y_2 están correlacionadas
4. Existen dos medidas por constructo

FIGURA 11A.1. Diagrama de secuencias de relaciones causales.

Construcción de las ecuaciones estructurales desde el diagrama de relaciones

El primer paso es transformar el diagrama de relaciones en una serie de ecuaciones estructurales para cada variable endógena. Las ecuaciones se muestran en la Tabla 11A.3.

TABLA 11 A.3. Ecuaciones del modelo estructural para el diagrama de relaciones

Variable endógena	*Variables exógenas*	*Variables endógenas*	*Error*
	X_1 X_2 X_3 X_4	Y_1 Y_2 Y_3	
Y_1 =	$b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$ +		+ ε_1
Y_2 =	$b_5X_1 + b_6X_2 + b_7X_3 + b_8X_4$ +	b_9Y_1	+ ε_2
Y_3 =		$b_{10}Y_1 + b_{11}Y_2$	+ ε_3

Notación de la correspondencia de los constructos e indicadores

Una vez que se han especificado las ecuaciones estructurales, se debe definir la medida de cada constructo. En este ejemplo, cada constructo tiene dos indicadores. La correspondencia de los indicadores y los constructos se muestra en la Tabla 11A.4.

Especificación de las ecuaciones del modelo estructural y de medida

Ahora podemos especificar el conjunto de ecuaciones, tanto para los modelos de medida, como para los estructurales. Sólo se incluyen aquellos coeficientes que se van a estimar. En primer lugar, definimos las ecuaciones de medidas y estructurales (véanse Tablas 11A.5 y 11A.6).

TABLA 11A.4. Indicadores y constructos del diagrama de relaciones en la notación LISREL

Exógeno		*Endógeno*	
Constructo	*Indicadores*	*Constructo*	*Indicadores*
X_1	V_1, V_2	Y_1	V_9, V_{10}
X_2	V_3, V_4	Y_2	V_{11}, V_{12}
X_3	V_5, V_6	Y_3	V_{13}, V_{14}
X_4	V_7, V_8		

TABLA 11 A.5. Ecuaciones del modelo de medida para el diagrama de relaciones

Indicador exógeno		*Constructos exógenos* ξ_1	ξ_2	ξ_3	ξ_4		*Error*
X_1	=	$\lambda^x_{11}\xi_1$				+	δ_1
X_2	=	$\lambda^x_{21}\xi_1$				+	δ_2
X_3	=		$\lambda^x_{32}\xi_2$			+	δ_3
X_4	=		$\lambda^x_{42}\xi_2$			+	δ_4
X_5	=			$\lambda^x_{53}\xi_3$		+	δ_5
X_6	=			$\lambda^x_{63}\xi_3$		+	δ_6
X_7	=				$\lambda^x_{74}\xi_4$	+	δ_7
X_8	=				$\lambda^x_{84}\xi_4$	+	δ_8

Indicador endógeno		*Constructos endógenos* η_1	η_2	η_3		*Error*
Y_1	=	$\lambda^y_{11}\eta_1$			+	ε_1
Y_2	=	$\lambda^y_{21}\eta_1$			+	ε_2
Y_3	=		$\lambda^y_{32}\eta_2$		+	ε_3
Y_4	=		$\lambda^y_{42}\eta_2$		+	ε_4
Y_5	=			$\lambda^y_{53}\eta_3$	+	ε_5
Y_6	=			$\lambda^y_{63}\eta_3$	+	ε_6

TABLA 11 A.6. Ecuaciones del modelo estructural para el diagrama de relaciones

Constructo endógeno		*Constructos exógenos* ξ_1 ξ_2 ξ_3 ξ_4		*Constructos endógenos* η_1 η_2 η_3		*Error*
η_1	=	$\gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_3 + \gamma_{14}\xi_4$	+		+	ζ_1
η_2	=	$\gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \gamma_{24}\xi_4$	+	$\beta_{21}\eta_1$	+	ζ_2
η_3	=			$\beta_{31}\eta_1 + \beta_{32}\eta_2$	+	ζ_3

El lector debería darse cuenta que las variables originales (V_1 a V_{14}) han de estar identificadas tanto como indicadores exógenos como endógenos. Las variables V_1 a V_8 corresponden de X_1 a X_8, respectivamente (los indicadores de constructos exógenos), y V_9 a V_{14} corresponden de Y_1 a Y_6 (los indicadores de constructos endógenos). Utilizando esta notación, presentamos las ecuaciones del modelo de medida en la Tabla 11A.6.

Especificación de las correlaciones de ecuaciones estructurales

A continuación, existen dos matrices de correlación relacionadas con las ecuaciones estructurales. La primera es la matriz phi, que denota las correlaciones entre los constructos exógenos. A partir del diagrama de relaciones, podemos ver que existen dos correlaciones de este tipo entre los constructos 1-2 (ϕ_{21}) y 3-4 (ϕ_{43}). Están anotadas en la matriz en la Tabla 11A.7. También existe una relación entre las ecuaciones estructurales para dos constructos endógenos (η_1 y η_3), representados en la matriz psi (ψ_{31}) en la Tabla 11A.7.

TABLA 11 A.7. Correlaciones entre constructos para el diagrama de relaciones

	Exógeno				*Endógeno*			
	ξ_1	ξ_2	ξ_3	ξ_4				
ξ_1	—				η_1	η_1	η_2	η_3
ξ_2	ϕ_{21}	—				—		
ξ_3			—		η_2		—	
ξ_4			ϕ_{43}	—	η_3	ψ_{31}		—

Correlaciones del modelo (indicador) de medida

La dos últimas matrices representan los errores de medida dentro del constructo entre los indicadores de los constructos endógenos y exógenos. Nótese que aunque las correlaciones puedan estar teóricamente entre cualquiera de los indicadores endógenos o exógenos, no pueden producirse entre indicadores de diferente tipo.

A efectos ilustrativos, se suponen dos medidas de las correlaciones de los errores. La primera está entre dos indicadores exógenos (X_1 y X_2), tal y como se muestra en la matriz theta-delta (Tabla 11A.8). La segunda correlación está entre dos indicadores endógenos (Y_3 e Y_4), que está representado en la matriz theta-epsilon (Tabla 11A.8).

TABLA 11 A.8. Correlación del error de medida para indicadores de constructos endógenos y exógenos del diagrama de relaciones

	Indicadores exógenos							
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	—							
X_2	$\phi_{\delta_{21}}$	—						
X_3			—					
X_4				—				
X_5					—			
X_6						—		
X_7							—	
X_8								—
	Indicadores endógenos							
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6		
Y_1	—							
Y_2		—						
Y_3			—					
Y_4			$\phi_{\varepsilon_{43}}$	—				
Y_5					—			
Y_6						—		

Hemos completado la especificación de las matrices LISREL necesarias para la estimación del modelo propuesto. La Figura 11A.2 ilustra el modelo al completo, con todas las relaciones representadas en un formato de modelo causal. En la representación gráfica de los modelos de medida y estructurales, se siguen algunas convenciones específicas. En primer lugar, como con los diagramas de relaciones, las flechas directas indican relaciones causales y las flechas curvadas representan correlaciones. En segundo lugar, los indicadores están representados por rectángulos y los constructos por elipses.

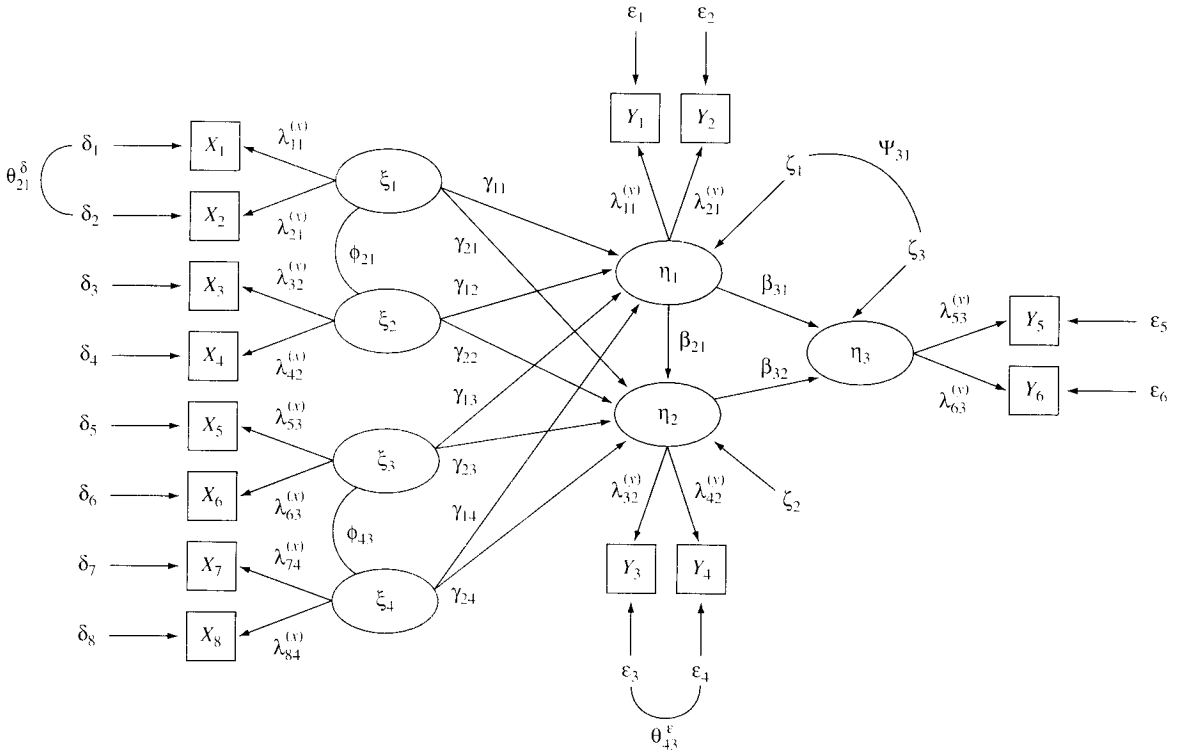


FIGURA 11A.2. Representación del diagrama completo del relaciones con la notación LISREL.

Resumen

La especificación de cualquier relación causal se puede incorporar directamente en una de las ocho matrices en la notación LISREL. Aunque no es necesario especificar los modelos en esta notación inicial, después de esta revisión de la forma de notación utilizada en SEM, los investigadores deberían sentirse preparados como para representar sus relaciones en su formato. Además, a medida que el investigador mejora su conocimiento de la notación, se pueden incorporar también relaciones más avanzadas.

APÉNDICE 11B

Medidas conjuntas de calidad del ajuste para modelos de ecuaciones estructurales

La evaluación de la bondad del ajuste conjunto para los modelos de ecuaciones estructurales no es tan directa como en otras técnicas de dependencia multivariante tales como la regresión múltiple, análisis discriminante, análisis multivariante de la varianza o incluso el análisis conjunto. SEM no tiene un único test estadístico que sea el que mejor describa la «fortaleza» de las predicciones del modelo. En su lugar, los investigadores han desarrollado varias medidas de bondad del ajuste que, cuando se usan en combinación, evalúan los resultados desde tres perspectivas: ajuste conjunto, ajuste comparativo respecto a un modelo base y la parsimonia del modelo. Las discusiones que se siguen presentan las medidas alternativas para cada una de estas perspectivas, junto con los métodos de cálculo de aquellas medidas que no están contenidas en los resultados que deben ser calculados por separado.

Surge una cuestión común en la discusión de cada medida: ¿cuál es un nivel de ajuste aceptable? Ninguna de estas medidas (excepto el estadístico chi-cuadrado) tiene un contraste estadístico asociado. Aunque en muchos casos se han sugerido líneas maestras, no tenemos ningún test absoluto y el investigador debe decidir en última instancia si el ajuste es aceptable. Bollen [22, pág. 275] trata directamente este tema: «En conjunto, la selección de un corte rígido de los índices de ajuste incremental es como seleccionar un R^2 mínimo de la ecuación de una regresión. Cualquier valor será controvertido. Tener conciencia de los factores que afectan a los valores y el buen juicio son las mejores guías para evaluar su tamaño». Este consejo se aplica igualmente bien a las medidas de calidad del ajuste.

Antes de examinar las diversas medidas del ajuste, puede ser útil revisar la derivación de los grados de libertad en los modelos estructurales. El número de valores de datos aislados en la matriz de entrada es s (donde $s = 1/2(k)(k-1)$ y k es el número total de indicadores tanto de los constructos endógenos como de los exógenos). Los grados de libertad (gl) para cualquier modelo estimado se calculan a continuación como $gl = s - t$, donde t es el número de coeficientes estima-

dos. Si el investigador sabe los grados de libertad para un modelo estimado y el número total de indicadores, entonces t puede calcularse directamente como $t=s\text{-}gl$.

El examen y obtención de las medidas de bondad del ajuste para SEM han suscitado un amplio interés entre los investigadores académicos en los últimos años, resultando en un continuo desarrollo de nuevas medidas de bondad del ajuste [36, 86, 90, 95, 100]. Esto se refleja en los programas estadísticos en la medida en que se modifican continuamente para proporcionar la información más relevante en relación con el modelo estimado. En nuestras discusiones hemos centrado nuestra atención en el programa LISREL, dada su amplia aplicación en las ciencias sociales. También este programa ha experimentado estos cambios. La nueva versión de LISREL ha aumentado sustancialmente el número y tipo de índices de ajuste directamente disponibles en el resultado. Por esta razón, la siguiente discusión y los datos del ejemplo detallan los cálculos de aquellas medidas no provistas en versiones anteriores del programa.

Medidas del ajuste absoluto

Las medidas del ajuste absoluto determinan el grado en que el modelo conjunto (modelos estructural y de medida) predice la matriz de correlación o covarianza observada. No se hace distinción en si el modelo de ajuste es mejor o peor en los modelos de medida o estructural. Entre las medidas del ajuste absoluto utilizadas habitualmente para evaluar SEM están los estadísticos de la chi-cuadrado, el parámetro de no centralidad, el estadístico de bondad del ajuste, la raíz cuadrada de la media de residuos al cuadrado, la raíz cuadrada de la media del error de aproximación al cuadrado y el índice de validación cruzada esperado.

Estadístico-ratio de verosimilitud chi-cuadrado

La medida más fundamental de ajuste conjunto es el estadístico-ratio de verosimilitud chi-cuadrado (χ^2), la única medida fundamentada estadísticamente de bondad del ajuste disponible en SEM [67]. Un gran valor de chi-cuadrado relativo a los grados de libertad significa que las matrices observadas y estimadas difieren considerablemente. Los niveles de significación estadística indican la probabilidad de que estas diferencias se deban solamente a variaciones de la muestra. Por tanto, valores de la chi-cuadrado bajos, que resultan en niveles de significación mayores que 0,05 o 0,01, indican que las matrices de entrada previstas y efectivas no son estadísticamente diferentes. En este caso, el investigador está buscando *diferencias no significativas* porque el contraste está entre las matrices actuales y las previstas. El investigador debe recordar que este método difiere de la necesidad habitual de encontrar significación estadística. Sin embargo, incluso la no significación estadística no garantiza el haber identificado el modelo «correcto», sino sólo que este modelo propuesto se ajusta bien a las varianzas y covarianzas observadas. No asegura al investigador que otro modelo no se ajustaría tan bien o mejor. El nivel de significación de 0,05 se recomienda como el mínimo aceptado, y los niveles de 0,1 y 0,2 deberían de ser superados antes de confirmar la no significación [42].

Una crítica importante de la medida de la chi-cuadrado es que es muy sensible a las diferencias del tamaño muestral, especialmente en los casos en que el tamaño muestral excede los 200 encuestados. A medida que el tamaño de muestra aumenta, esta medida tiene una gran tendencia a indicar diferencias significativas para modelos equivalentes. Si el tamaño muestral se hace suficientemente grande, se encontrarían diferencias significativas para cualquier modelo especificado. Además, a medida que el tamaño muestral se acerca a 100 o incluso menos, el test de la chi-cuadrado mostrará un ajuste aceptable (diferencias no significativas en las matrices de entrada previstas y observadas), incluso cuando ninguna de las relaciones del modelo se muestren estadísticamente significativas. Por tanto, el estadístico de la chi-cuadrado es muy sensible de diversas maneras tanto a los tamaños de muestra pequeños como a los grandes, y se anima al in-

vestigador a complementar esta medida con otras medidas de ajuste en todos los casos. El uso de la chi-cuadrado es apropiado para tamaños de muestra entre 100 y 200, haciéndose menos significativo, menos fiable, cuando los tamaños de muestra caigan fuera de este rango.

La sensibilidad de la medida de chi-cuadrado se extiende a las consideraciones anteriores del tamaño muestral. Por ejemplo, se ha mostrado que esta medida cambia basándose en el número de categorías en la variable de respuesta [52]. Dada su sensibilidad estadística a muchos factores, se anima al investigador a complementar la medida de la chi-cuadrado con otras medidas de calidad del ajuste.

Parámetros de no centralidad y no centralidad a escala

El parámetro de no centralidad (NCP) es el resultado de la búsqueda de los estadísticos de una medida alternativa al ratio-estadístico de verosimilitud chi-cuadrado que se vea menos afectada por el tamaño muestral. La teoría estadística sugiere que una medida chi-cuadrado de no centralidad que se verá menos afectada por el tamaño muestral en su representación de las diferencias entre las matrices de datos estimados y efectivos [76]. En un problema LISREL, el parámetro de no centralidad se puede calcular como sigue:

$$\text{NCP} = \chi^2 - \text{Grados de libertad}$$

Aunque esta medida ajusta la χ^2 por los grados de libertad del modelo estimado, todavía está en términos del tamaño de muestra original. Para «estandarizar» la NCP, se divide por el tamaño de muestra para obtener en parámetro de no centralidad a escala (SNCP) [76]. Éste se puede calcular como:

$$\text{SNCP} = \frac{\chi^2 - \text{Grados de libertad}}{\text{Tamaño de la muestra}}$$

Esta medida ajustada es análoga a la medida de la distancia media euclídea al cuadrado entre el modelo estimado y el modelo sin restricción [76]. Tanto para los parámetros a escala como para los que no lo están, el objetivo es minimizar el valor del parámetro. Dado que no existe ningún test estadístico para esta medida, se utiliza mejor para hacer comparaciones entre modelos alternativos.

Índice de bondad del ajuste

El índice de bondad del ajuste (GFI) [64, 67] es otra medida ofrecida por LISREL. Es una medida no estadística que va en valor desde 0 (mal ajuste) a 1,0 (ajuste perfecto). Representa el grado de ajuste conjunto (los residuos al cuadrado de la predicción comparado con los datos efectivos), pero no está ajustada por los grados de libertad. Altos valores indican un mejor ajuste, pero no se ha establecido ningún umbral absoluto de aceptabilidad.

Residuo cuadrático medio

El residuo cuadrático medio (RMSR)es la raíz cuadrada de la medida de los residuos al cuadrado –una media de los residuos entre las matrices de entrada estimadas y observadas. Si se utilizan las covarianzas, el RMSR es la media de los residuos de la covarianza. Si se utiliza un matriz de correlación, entonces el RMSR está en términos de una correlación residual media. La RMSR es más útil para correlaciones, que están todas a la misma escala, que para covarianzas, que pueden diferir de variable a variable dependiendo de la unidad de medida. De nuevo, no se puede establecer ningún nivel

umbral, pero el investigador puede evaluar la significación práctica de la magnitud de RMSR a la luz de los objetivos de investigación y de las correlaciones o covarianzas efectivas y observadas.

Error de aproximación cuadrático medio

Otra medida que intenta corregir la tendencia del estadístico chi-cuadrado para rechazar cualquier modelo especificado con una muestra suficientemente grande es el error de aproximación cuadrático medio (RMSEA). Al igual que RMSR, el RMSEA es la discrepancia por grado de libertad. Difiere de RMSR, sin embargo, en que la discrepancia se mide en términos de la población, no sólo en términos de la muestra utilizada para la estimación [93]. El valor es representativo de la bondad del ajuste que podría esperarse si el modelo fuera estimado con la población, no sólo con la muestra extraída de la estimación. Los valores que van de 0,05 a 0,08 se consideran aceptables. Un examen empírico de diversas medidas encontró que el RMSEA se ajustaba mejor para su uso en la estrategia confirmatoria o de modelos rivales con grandes muestras [86].

Índice de validación cruzada esperada

El índice de validación cruzada esperada (ECVI) es una aproximación a la bondad del ajuste que conseguiría el modelo estimado en otra muestra del mismo tamaño. Basándose en la matriz de covarianza de la muestra, tiene en cuenta el tamaño de muestra efectivo y la diferencia que podría esperarse en otra muestra. El ECVI también tiene en cuenta el número de parámetros estimados tanto para los modelos de medida como para los estructurales. El ECVI se calcula como :

$$\text{ECVI} = \frac{\chi^2}{\text{Tamaño de la muestra} - 1} + \frac{2 \times \text{número de parámetros estimados}}{\text{Tamaño de la muestra} - 1}$$

El EVCI no tiene un rango especificado aceptable de valores, pero se utiliza en la comparación entre modelos alternativos.

Índice de validación cruzada

El índice de validación cruzada (CVI) evalúa la calidad del ajuste cuando se ha llevado a cabo una validación cruzada. La validación cruzada se desarrolla en dos pasos. En primer lugar, la muestra conjunta se divide en dos muestras –una muestra de estimación muestra y otra de validación. La muestra de estimación se utiliza para estimar un modelo y crear la correlación estimada de la matriz de covarianza. Esta matriz se compara a continuación con la muestra de la validación de la muestra. Se puede desarrollar un proceso de validación cruzada doble comparando la matriz de correlación o covarianza estimada de cada muestra con una matriz de datos de otra muestra.

Medidas del ajuste incremental

La segunda clase de medidas compara el modelo propuesto con algún modelo de referencia, más a menudo referido como el modelo nulo. El modelo nulo debería ser un modelo realista que se espera que el resto de los modelos sobrepasen. En la mayoría de los casos, el modelo nulo es un modelo de constructo único con todos los indicadores midiendo perfectamente el constructo (es decir, esto representa el valor de la chi-cuadrado asociada con la varianza total en el conjunto de correlaciones o covarianzas). Existe, sin embargo, cierto desacuerdo sobre cómo especificar con exactitud el modelo nulo en muchas situaciones [91].

Índice ajustado de bondad del ajuste

Este índice (AGFI) es una extensión de GFI, ajustado por el ratio entre los grados de libertad del modelo propuesto y los grados de libertad del modelo nulo. Es muy parecido al índice de ajuste de parsimonia normada (discutido después), y un nivel aceptable y recomendado es un valor mayor o igual a 0,90.

Índice Tucker-Lewis

La siguiente medida de ajuste incremental es el índice de Tucker-Lewis [97], también conocido como el índice de ajuste no normado (NNFI). Primeramente propuesto como medio de evaluar el análisis factorial, el TLI se ha extendido a SEM. Combina una medida de parsimonia en un índice comparativo entre los modelos nulos y propuestos, lo que resulta en valores que van de 0 a 1,0. Se expresa como:

$$\text{TLI} = \frac{(\chi^2_{\text{nulo}} / gl_{\text{nulo}}) - (\chi^2_{\text{propuesto}} / gl_{\text{propuesto}})}{(\chi^2_{\text{nulo}} / gl_{\text{nulo}}) - 1}$$

Un valor recomendado para TLI es un nivel de 0,90 o superior. Esta medida también se puede utilizar para comparar entre modelos alternativos mediante la sustitución del modelo alternativo por el modelo nulo.

Índice del ajuste normal

Una de las medidas más populares es el índice de ajuste normado (NFI) [16], que es una medida que va de 0 (ningún ajuste) a 1,0 (ajuste perfecto). De nuevo, el NFI es una comparación relativa del modelo propuesto al modelo nulo. El NFI se calcula como:

$$\text{NFI} = \frac{(\chi^2_{\text{nulo}} - \chi^2_{\text{propuesto}})}{\chi^2_{\text{nulo}}}$$

Como con el índice Tucker-Lewis, no existe un valor absoluto que indique un nivel de ajuste aceptable, pero un valor recomendado normalmente es 0,90 o superior.

Otras medidas de ajuste incremental

Se han propuesto varias medidas de ajuste incremental, y la nueva versión de LISREL incluye tres en su paquete. El índice de ajuste relativo (RFI), el índice de ajuste incremental (IFI) y el índice de ajuste comparado (CFI) representan todos comparaciones entre el modelo estimado y el modelo nulo o independiente. Los valores caen entre 0 y 1,0, indicando los valores elevados altos niveles de calidad del ajuste. El CFI se ha encontrado más apropiado en una estrategia de desarrollo del modelo o cuando está disponible un tamaño de muestra menor [86]. El lector interesado puede encontrar detalles específicos de cada medida en lecturas seleccionadas [15, 21, 22].

Medidas de ajuste de parsimonia

Las medidas de ajuste de parsimonia relacionan la calidad del ajuste del modelo al número de coeficientes estimados exigidos para conseguir este nivel de ajuste. Sus objetivos básicos son diag-

nosticar si el ajuste del modelo se ha conseguido debido a un «sobreajuste» de los datos por tener demasiados coeficientes. Este procedimiento es similar al «ajuste» del R^2 en una regresión múltiple. Sin embargo, dado que no existe ningún test de significación estadístico, su uso en un sentido absoluto está limitado en muchos casos de comparaciones entre modelos.

Índice de ajuste normado de parsimonia

La primera medida de ajuste en este caso es el índice de ajuste normado de parsimonia (PNFI) [61], una modificación del NFI. El PNFI tiene en cuenta el número de grados de libertad utilizados para conseguir un nivel de ajuste. La parsimonia se define en la consecución de mayores niveles de ajuste por grado de libertad utilizado (un grado de libertad por coeficiente estimado). Por tanto, es deseable la mayor parsimonia. La PNFI se define como:

$$\text{PNFI} = \frac{gl_{\text{propuesto}}}{gl_{\text{nulo}}} \times \text{NFI}$$

Valores elevados de PNFI son mejores, y su uso principal consiste en la comparación de modelos con diferentes grados de libertad. Se utiliza para comparar modelos alternativos y no existen niveles recomendados de ajuste. Sin embargo, cuando se compara entre modelos, se proponen diferencias de 0,06 a 0,09, que sean indicativas de diferencias sustanciales en los modelos [103].

Índices de calidad de ajuste de parsimonia

El índice de calidad de ajuste de parsimonia (PGFI) modifica el GFI de forma diferente del AGFI. Mientras que el ajuste del AGFI del GFI se basaba en los grados de libertad en los modelos estimados y nulos, el PGFI se basa en la parsimonia del modelo estimado. Ajusta el GFI de la siguiente forma:

$$\text{PGFI} = \frac{gl_{\text{propuesto}}}{1/2\ (\text{Núm. de variables manifiestas})\ (\text{Núm. de variables manifiestas} + 1)} \times \text{GFI}$$

El valor varía entre 0 y 1,0, con valores elevados indicando una mayor parsimonia del modelo.

Chi-cuadrado normada

Jöreskog [63] propuso que la chi-cuadrado se «ajustase» por los grados de libertad para evaluar el ajuste de modelo en varios modelos. Esta medida puede denominarse la chi-cuadrado normada, y es el ratio de la chi-cuadrado dividida por los grados de libertad. Esta medida ofrece dos formas de evaluar los modelos inapropiados: (1) un modelo que puede estar «sobreajustado», por tanto su baja significación se debe a la casualidad, tipificado por valores menores que 1,0; y (2) modelos que no son verdaderamente representativos de los datos observados y por tanto, su mejora es necesaria teniendo valores mayores que un umbral superior, tanto 2,0 o 3,0 [30] o un límite más liberal de 5,0 [63]. Sin embargo, dado que el valor de la chi-cuadrado es el principal componente de esta medida, no está sujeto a efectos de tamaño muestral discutidos previamente con relación al estadístico de la chi-cuadrado.

La chi-cuadrado normada, sin embargo, se ha mostrado que es, de alguna forma, poco fiable [56, 100], de tal forma que los investigadores siempre la combinan con otras medidas de bondad del ajuste.

Criterio de información de Akaike

Otra medida basada en la teoría de la información estadística es el criterio de información de Akaike [1]. Al igual que PNFI, el AIC es una medida comparativa entre modelos con diferente número de constructos. El AIC se calcula como:

$$\text{AIC} = \chi^2 + 2 \times \text{número de parámetros estimados}$$

Los valores AIC cercanos a cero indican un mejor ajuste y una mayor parsimonia. Un valor reducido de AIC se produce cuando los valores de la chi-cuadrado se consiguen con unos pocos coeficientes estimados. Con ello obtenemos no sólo un buen ajuste de las covarianzas y correlaciones previstas y observadas, sino también un modelo no propenso al «sobreajuste».

Una revisión de las medidas de calidad del ajuste estructural

La Tabla 11B.1 presenta un resumen de las medidas de calidad del ajuste para el modelo estructural estimado en el Capítulo 11. En esta tabla se demuestra la forma de obtención y la interpretación de cada medida. Como ilustra este ejemplo, el investigador busca diversas medidas para evaluar diferentes aspectos del modelo, y se espera que todas las medidas indiquen acuerdo sobre el nivel de aceptabilidad del modelo.

Al evaluar el conjunto de medidas, se aplican algunos criterios generales y se indican modelos con un ajuste aceptable.

- χ^2 no significativos (al menos $\rho > 0{,}005$, quizá 0,10 o 0,20).
- Índices de ajuste incremental (NFI, TLI) mayores que 0,90.
- Bajos valores de RMSR y RMSEA basados en el uso de covarianzas y correlaciones.
- Índices de parsimonia que representan el modelo propuesto como más parsimonioso que los modelos alternativos.

Como se ha indicado previamente, el investigador debería evaluar los modelos propuestos sobre una serie de medidas de cada tipo. Se debería alcanzar consenso sobre la aceptabilidad del modelo sólo después de un examen de los resultados del conjunto completo de medidas de calidad del ajuste.

Resumen

Los tipos y variedad de medidas de calidad del ajuste están aumentado a medida que los investigadores continúan explorando las posibilidades de la modelización de ecuaciones estructurales. El usuario, sin embargo, se enfrenta con la tarea de no sólo seleccionar las medidas apropiadas sino también evaluar por patrones claramente subjetivos si el modelo es aceptable. El efecto final es la incertidumbre acerca de si es aceptable frente a inaceptable, dejando el peso de la demostración al investigador, en lugar de que éste confíe en un contraste estadístico.

El pensamiento prevaleciente mantiene que el test más fuerte de cualquier modelo propuesto se consigue a través de la comparación del modelo con cualquier modelo propuesto. Las medidas de ajuste incremental fueron un anticipo para ofrecer un modelo alternativo (estándar), denominado el modelo nulo o independiente. Pero la incapacidad de LISREL o cualquier otro programa SEM para asegurar que ningún otro modelo tendrá un mejor ajuste a los datos que el modelo propues-

TABLA 11B.1. Comparación de las medidas de bondad del ajuste para el modelo estructural

Datos del modelo estructural:
Quince indicadores para cinco constructos (tres exógenos, dos endógenos)
Total grados de libertad: $^1/_2(15 \times 16) = 120$
Número de parámetros estimados (modelos de medida y estructural): 35
Tamaño de muestra: 136
Modelo propuesto: $\chi^2 = 178{,}714$ $gl = 85$ $p = 0{,}000$
Modelo nulo e independiente: $\chi^2 = 1.040{,}194$ $gl = 105$ $p = 0{,}000$

Evaluación del modelo estructural con medidas de bondad del ajuste

Medida de bondad del ajuste	*Niveles de ajuste aceptable*	*Medida del cálculo*	*Aceptabilidad*[a]
Medidas del ajuste absoluto			
Estadístico ratio de verosimilitud chi-cuadrado (χ^2)	Se ofrecen test estadísticos de significación	$\chi^2 = 178{,}714$[b] significance level: 000	Marginal
Parámetro de no centralidad (NCP)	Fijado en términos de la χ^2, juzgada en comparación con modelos alternativos	NCP = 93,714	No aplicable
Parámetro de no centralidad a escala (SNCP)	NCP fijado en términos de diferencia media por observación para comparación entre modelos	$\text{SNCP} = \frac{93{,}714}{136} = 0{,}689$	No aplicable
Índice de bondad del ajuste (GFI)	Valores elevados indican mejores ajustes, no hay umbrales establecidos	GFI = 0,865[b]	Marginal
Residuo cuadrático medio (RMSR)	Fijado en términos de la matriz de entrada (covarianza o correlación), con niveles aceptables fijados por el analista	RMSR = 0,076[b]	Marginal
Error de aproximación cuadrático medio (RMSEA)	Diferencia media por grado de libertad esperado que se produzca en la población, no en la muestra. Niveles aceptables por debajo de 0,08	RMSEA = 0,090	Marginal
Índice de validación cruzada esperada (ECVI)	La bondad del ajuste esperada en otra muestra del mismo tamaño. No hay rango establecido de valores aceptables. Utilizado en la comparación entre modelos	$\text{ECVI} = \frac{178{,}714}{136-1} + \frac{2 \times 35}{136-1} = 1{,}842$	No aplicable
Medidas de ajuste incremental			
Índice Tucker–Lewis (TLI) o NNFI	Nivel recomendado: 90	$\text{TLI} = \frac{(1.040{,}194/105) - (178{,}714/85)}{(1.040{,}194/105) - 1} = 0{,}876$	Marginal
Índice de ajuste normado (NFI)	Nivel recomendado: 90	$\text{NFI} = \frac{1.040{,}194 - 178{,}714}{(1.040{,}194)} = 0{,}828$	Marginal

TABLA 11B.1. (Continuación)

Índice ajustado de bondad del ajuste (AGFI)	Nivel recomendado: 90	AGFI = 0,810[b]	Marginal
Medidas de ajuste de parsimonia			
Índice de bondad de ajuste de parsimonia (PGFI)	Una reespecificación del GFI con altos valores que reflejan una mayor parsimonia del modelo. Utilizada para la comparación entre modelos	$PGFI = \frac{85}{1/2\,(15 \times 16)} \times 0{,}865 = 0{,}613$	No aplicable
Chi-cuadrado normada	Nivel recomendado: Límite inferior: 1,0 Límite superior: 2,0/3,0 o 5,0	Normado $\chi^2 = \frac{178{,}714}{85} = 2{,}103$	Marginal
Índice de ajuste normado de parsinomia	Valores elvados que indican un menor ajuste, utilizado sólo en la comparación entre modelos alternativos	$PNFI = \frac{85}{105} \times 0{,}828 = 0{,}670$	No aplicable
Criterio de información de Akaike (AIC)	Valores positivos pequeños indican parsinomia, utilizada en comparación de modelos alternativos	AIC = 178,714 + 2 (35) = 248,714	No aplicable

[a] Un nivel de aceptabilidad de «no aplicable» es aplicado para medidas usadas sólo en compración con modelos alternativos.
[b] Las medidas provienen, en todas las versiones, de LISREL.

to hace del proceso de comparación formalizada entre los modelos alternativos o rivales el test más estricto de la teoría. Se anima a los investigadores a examinar los modelos alternativos no en un enfoque de desarrollo del modelo, sino como contraste de modelos alternativos para encontrar la mejor representación del modelo teórico propuesto.

Bibliografía

1. Akaike, H. (1987), «Factor Analysis and AIC.» *Psychometrika* 52: 317-32.
2. Allison, Paul D. (1987), «Estimation of Linear Models with Incomplete Data.» In C. Clogg (ed.) *Sociological Methodology*, San Francisco: Jossey-Bass, pp. 71-103.
3. Anderson J. C., and D. W. Gerbing (1982), «Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measures.» *Journal of Marketing Research* 19 (November): 453-60.
4. Anderson, J. C., and D. W. Gerbing (1988), «Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach.» *Psychological Bulletin* 103: 411-23.
5. Anderson, J. C., and D. W. Gerbing (1990), «Assumptions and Comparative Strengths of the Two-Step Approach: Comment on Fornell and Yi.» *Sociological Methods and Research* 20: 321-33.
6. Arbuckle, J. L. (1994), «AMOS—Analysis of Moment Structures,» *Psychometrika* 59(1): 135-37.
7. Austin, J. T., and R. F. Calderon (1996), «Theoretical and Technical Contributions to Structural Equation Modeling: An Updated Annotated Bibliography,» *Structural Equation Modeling* 3(2): 105-25.
8. Bagozzi, R. P. (1980), *Causal Models in Marketing.* New York: Wiley.
9. Bagozzi, R. P., and Y. Yi (1988), «On the Use of Structural Equation Models in Experimental Designs.» *Journal of Marketing Research* 26 (August): 271-84.
10. Becker, Paul A., Arjen Merckens, and Toms J. Wansbrek (1994), *Identification, Equivalent Models and Computer Algebra.* New York: Academic Press.
11. Bentler, P. M. (1980), «Multivariate Analysis with Latent Variables: Causal Modeling.» *Annual Review of Psychology* 31: 419-56.
12. Bentler, P. M. (1986), «Structural Modeling and Psychometrika: A Historical Perspective on Growth and Achievements.» *Psychometrika* 51: 35-51.
13. Bentler, P. M. (1988), *Theory and Implementation of*

EQS, a Structural Equations Program, Los Angeles: BMDP Statistical Software.
14. Bentler, P. M. (1992), *EQS: Structural Equations Program Manual.* Los Angeles: BMDP Statistical Software.
15. Bentler, P. M. (1990), «Comparative Fit Indexes in Structural Models.» *Psychological Bulletin* 107: 238-46.
16. Bentler, P. M., and D. G. Bonnett. (1980), «Significance Tests and Goodness of Fit in the Analysis of Covariance Structures.» *Psychological Bulletin* 88: 588-606.
17. Bentler, P. M., and C. Chou (1987), «Practical Issues in Structural Modeling.» *Sociological Methods and Research* 16 (August): 78-117.
18. Bentler, P. M., and E. J. C. Wu (1993), *EQS/Windows User's Guide.* Los Angeles: BMDP Statistical Software.
19. Blalock, H. M. (1982), *Conceptualization and Measurement in the Social Sciences.* Beverly Hills, Calif.: Sage.
20. Blalock, H. M. (1985), *Causal Modeling in the Social Sciences.* New York: Academic Press.
21. Bollen, K. A. (1986), «Sample Size and Bentler and Bonnett's Nonnormed Fit Index.» *Psychometrika* 51: 375-77.
22. Bollen, K. A. (1989), *Structural Equations with Latent Variables.* New York: Wiley.
23. Bollen, K. A., and K. G. Joreskog (1985), «Uniqueness Does Not Imply Identification.» *Sociological Methods and Research* 14: 155-63.
24. Bollen, K. A., and R. A. Stine (1993), «Bootstrapping Goodness-of-Fit Measures in Structural Equation Models.» In K. A. Bollen and J. S. Long (eds.), *Testing Structural Equation Models*, Newbury Park, Calif.: Sage.
25. Breckler, S. J. (1990), «Applications of Covariance Structure Modeling in Psychology: Cause for Concern?» *Psychological Bulletin* 107(2): 260-73.
26. Brown, R. L. (1994), «Efficacy of the Indirect Approach for Estimating Structural Equation Models with Missing Data: A Comparison of Five Methods,» *Structural Equation Modeling* 1(4): 375-80.
27. Bullock, H. E., L. L. Harlow, and S. Mulaik (1994), «Causation Issues in Structural Equation Modeling.» *Structural Equation Modeling* 1(3): 253-67.
28. Burt, R. S. (1976), «Interpretational Confounding of Unidimensional Variables in Structural Equation Modeling.» *Sociological Methods and Research* 5: 3-51.
29. Byrne, B. M. (1995), «Strategies in Testing for an Invariant Second-Order Factor Structure: A Comparison of EQS and LISREL,» *Structural Equation Modeling* 2(1): 53-72.
30. Carmines, E., and J. McIver (1981), «Analyzing Models with Unobserved Variables: Analysis of Covariance Structures.» In G. Bohrnstedt and E. Borgatta (eds.), *Social Measurement: Current Issues*, Beverly Hills, Calif.: Sage.
31. Costner, H. L., and R. Schoenberg (1979), «Diagnosing Indicator Ills in Multiple Indicator Models.» In A. Goldberger and O. Duncan (eds.), *Structural Equation Models in the Social Sciences*, New York: Seminar Press.
32. Cronbach, L. J. (1951), «Coefficient Alpha and the Internal Structure of Tests.» *Psychometrica* 16: 297-334.
33. Darden, W. R. (1981), «Review of Behavioral Modeling in Marketing.» In Ben M. Enis and Kenneth J. Roering (eds.), *Review of Marketing.* Chicago: American Marketing Association.
34. Davis, W. R. (1993), «The FCI Rule of Identification for Confirmatory Factor Analysis: A General Sufficient Condition.» *Sociological Methods and Research* 21: 403-37.
35. Dillon, W., A. Kumar, and N. Mulani (1987), «Offending Estimates in Covariance Structure Analysis—Comments on the Causes and Solutions to Heywood Cases.» *Psychological Bulletin* 101: 126-35.
36. Ding, L., W. F. Velicer, and L. L. Harlow (1995), «Effects of Estimation Methods, Number of Indicators per Factor and Improper Solutions on Structural Equation Modeling Fit Indices.» *Structural Equation Modeling*, 2: 119-43.
37. Dolan, C. (1996), «Principal Component Analysis Using LISREL 8,» *Structural Equation Modeling* 3(4): 307-22.
38. Duncan, O. D. (1966), «Path Analysis: Sociological Examples.» *American Journal of Sociology* 72: 1-16.
39. Duncan, O. D. (1975), *Introduction to Structural Equation Models.* New York: Academic Press.
40. Fan, Xitao (1997), «Canonical Correlation Analysis and Structural Equation Modeling: What Do They Have in Common?» *Structural Equation Modeling* 4(1): 65-79.
41. Fassinger, R. E. (1987), «Use of Structural Equation Modeling in Counseling Psychology Research.» *Journal of Counseling Psychology* 34: 425-36.
42. Fornell, C. (1983), «Issues in the Application of Covariance Structure Analysis: A Comment.» *Journal of Consumer Research* 9: 443-48.
43. Fornell, C., and D. F. Larker (1981), «Evaluating Structural Equation Models with Unobservable Variables and Measurement Error.» *Journal of Marketing Research* 18 (February): 39-50.
44. Fornell, C., and Y. Yi (1992), «Assumptions of the Two-Step Approach to Latent Variable Modeling.» *Sociological Methods and Research* 20: 291-320.
45. Fornell, C., and Y. Yi (1992), «Assumptions of the Two-Step Approach: Reply to Anderson and Gerbing.» *Sociological Methods and Research* 20: 334-39.
46. Fraser, C. (1980), *COSAN User's Guide.* Toronto: Ontario Institute for Studies in Education.
47. Gerbing, D. W., and J. C. Anderson (1984), «On the Meaning of Within-Factor Correlated Measurement Errors.» *Journal of Consumer Research* 11: 572-80.
48. Gerbing, D. W., and J. G. Hamilton (1996), «Viability

of Exploratory Factor Analysis as a Precursor to Confirmatory Factor Analysis,» *Structural Equation Modeling* 3(1): 62-72.
49. Glymour, C. (1988), *Discovering Causal Structure.* Orlando, Fla.: Academic Press.
50. Glymour, C., R. Scheines, P. Spirtes, and K. Kelly (1987), «Discovering Casual Structure: Artificial Intelligence.» *Philosophy of Science and Statistical Models*, New York: Academic Press.
51. Goldberger, A. S., and O. D. Duncan (1973), *Structural Equation Models in the Social Sciences.* New York: Seminar Press.
52. Green, S. B., T. M. Akey, K. K. Fleming, S. C. Hershberger, and J. G. Marquis (1997), «Effect of the Number of Scale Points on Chi-Square Fit Indices in Confirmatory Factor Analysis.» *Structural Equation Modeling* 4(2): 108-20.
53. Hatcher, L. (1996), «Using SAS® PROC CALIS for Path Analysis: An Introduction.» *Structural Equation Modeling* 3(2): 176-92.
54. Hayduk, L. A. (1987), *Structural Equation Modeling with LISREL: Essentials and Advances.* Baltimore: Johns Hopkins University Press.
55. Hayduk, L. A. (1996), *LISREL Issues, Debates and Strategies.* Baltimore: Johns Hopkins University Press.
56. Hayduk, L. A. (1987), *Structural Equation Modeling with LISREL.* Baltimore: Johns Hopkins University Press.
57. Heise, D. R. (1975), *Causal Analysis.* New York: Wiley.
58. Hoelter, J. W. (1983), «The Analysis of Covariance Structures: Goodness-of-Fit Indices.» *Sociological Methods and Research* 11: 325-44.
59. Hox, J. J. (1995), «AMOS, EQS and LISREL for Windows: A Comparative Review.» *Structural Equation Modeling* 2(1): 79-91.
60. Hunt, S. D. (1990), *Marketing Theory: The Philosophy of Marketing Science.* Homewood, Ill.: Irwin.
61. James, L. R., S. A. Muliak, and J. M. Brett (1982), *Causal Analysis: Assumptions, Models and Data.* Beverly Hills, Calif.: Sage.
62. Jöreskog, K. G. (1969), «A General Approach to Confirmatory Maximum Likelihood Factor Analysis.» *Psychometrika* 34: 183-202.
63. Jöreskog, K. G. (1970), «A General Method for Analysis of Covariance Structures.» *Biometrika* 57: 239-51.
64. Jöreskog, K. G., and D. Sorbom (1988), *LISREL VII: Analysis of Linear Structure Relationships by the Method of Maximum Likelihood.* Mooresville, Ill.: Scientific Software.
65. Jöreskog, K. G., and D. Sorbom (1988), *LISREL 7.* Chicago: SPSS, Inc.
66. Jöreskog, K. G., and D. Sorbom (1993), *PRELIS2: A Program for Multivariate Data Screening and Data Summarization.* Mooresville, Ill.: Scientific Software.
67. Jöreskog, K. G., and D. Sorbom (1993), *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language.* Mooresville, Ill.: Scientific Software.
68. Kenny, D. A. (1979), *Correlation and Causation.* New York: Wiley.
69. Lee, S., and S. Hershberger (1990), «A Simple Rule for Generating Equivalent Models in Structural Equation Modeling.» *Multivariate Behavioral Research* 25: 313-34.
70. Loehlin, J. C. (1987), *Latent Variable Models: An Introduction to Factor, Path and Structural Analysis.* Hillsdale, N.J.: Lawrence Erlbaum.
71. Long, J. S. (1983), *Covariance Structure Models: An Introduction to LISREL.* Beverly Hills, Calif.: Sage.
72. MacCullum, R. (1986), «Specification Searches in Covariance Structure Modeling.» *Psychological Bulletin* 100: 107-20.
73. Marsh, H. W., J. R. Balla, and R. P. McDonald (1988), «Goodness-of-Fit Indices in Confirmatory Factor Analysis: The Effect of Sample Size.» *Psychological Bulletin* 103: 391-410.
74. Marsh, H. W., and D. Hoceuar (1985), «Application of Confirmatory Factor Analysis to the Study of Self-Concept: First- and Higher-Order Factor Models and Their Invariance Across Groups.» *Psychological Bulletin* 97(1): 562-82.
75. Marsh, H. W., and D. Hoceuar (1994), «Confirmatory Factor Analysis Models of Factorial Invariance: A Multifaceted Approach.» *Structural Equation Modeling* 1(10): 5-34.
76. McDonald, R. P., and H. W. Marsh (1990), «Choosing a Multivariate Model: Noncentrality and Goodness of Fit.» *Psychological Bulletin* 107: 247-55.
77. Mulaik, S. A., L. R. James, J. Van Alstine, N. Bennett, S. Lind, and D. C. Stillwell (1989), «An Evaluation of Goodness of Fit Indices for Structural Equation Models.» *Psychological Bulletin* 103: 430-55.
78. Mulaik, S. A., Larry R. James, Judith Van Alstine, Nathan Bennett, Sherri Lind, and C. Dean Stilwell (1989), «Evaluation of Goodness-of-Fit Indices for Structural Equation Models.» *Psychological Bulletin* 105: 430-45.
79. Muthen, B., and D. Kaplan (1985), «A Comparison of Methodologies for the Factor Analysis of Nonnormal Likert Variables.» *British Journal of Mathematical and Statistical Psychology* 38: 171-89.
80. Neale, M. C., A. C. Heath, J. K. Kewitt, L. J. Eaves, and D. W. Walker (1989), «Fitting Genetic Models with LISREL: Hypothesis Testing.» *Behavior Genetics* 19: 37-49.
81. O'Brien, R. M., and T. Reilly (1995), «Equality in Constraints and Metric-Setting Measurement Models.» *Structural Equation Modeling* 2(1): 53-72.
82. Predhazur, E. J., and L. P. Schmelkin (1992), *Measurement Design and Analysis: An Integrated Approach.* Hillsdale, N.J.: Lawrence Erlbaum and Associates.
83. Raykou, T., and K. F. Widaman (1995), «Issues in Applied Structural Equation Modeling Research.» *Structural Equation Modeling* 2(4): 289-318.
84. Reilly, T. (1995), «A Necessary and Sufficient Condition for Identification of Confirmatory Factor Ana-

lysis Models of Complexity One.» *Sociological Methods and Research* 23(4): 421-41.
85. Rigdon, E. E. (1995), «A Necessary and Sufficient Identification Rule for Structural Models Estimated in Practice.» *Multivariate Behavior Research* 30(3): 359-83.
86. Rigdon, E. E. (1996), «CFI versus RMSEA: A Comparison of Two Fit Indices for Structural Equation Modeling.» *Structural Equation Modeling* 3(4): 369-79.
87. Rigdon, E. E. (1994), «Demonstrating the Effects of Unmodeled Random Measurement Error.» *Structural Equation Modeling* 1(4): 375-80.
88. Robles, J. (1996), «Confirmation Bias in Structural Equation Modeling.» *Structural Equation Modeling* 3(1): 307-22.
89. Rubio, D. M., and D. E. Gillespie (1995), «Problems with Error in Structural Equation Models», *Structural Equation Modeling* 2(4): 367-78.
90. Satorra, A., and P. Bentler (1994), «Correction to Test Statistics and Standard Errors in Covariance Structure Analysis. In A. Von Eye and C. Clogg (eds.), Latent Variable Analysis: Applications for Developmental Research, Newbury Park, Calif.: Sage, pp. 399-419.
91. Sobel, M. E., and G. W. Bohrnstedt (1985), «The Use of Null Models in Evaluating the Fit of Covariance Structure Models.» In Nancy Brandon Tuma (ed.), *Sociological Methodology*, San Francisco: Jossey-Bass.
92. Steenkamp, J. E. M., and H. C. M. van Trijp (1991), «The Use of LISREL in Validating Marketing Constructs.» *International Journal of Research in Marketing* 8(4): 283-99.
93. Steiger, J. H. (1990), «Structural Model Evaluation and Modification: An Interval Estimation Approach.» *Multivariate Behavioral Research* 25: 173-80.
94. Tanaka, J. (1987), «How Big Is Enough? Sample Size and Goodness-of-Fit in Structural Equation Models with Latent Variables.» *Child Development* 58: 134-46.
95. Tanaka, J. (1993), «Multifaceted Conceptions of Fit in Structural Equation Models.» In K. A. Bollen and J. S. Long (eds.), *Testing Structural Equation Models*, Newbury Park, Calif.: Sage.
96. Tremblay, P. F., and R. G. Gardner (1996), «On the Growth of Structural Equation Modeling in Psychological Journals.» *Structural Equation Modeling* 3(2): 93-104.
97. Tucker, L. R., and C. Lewis (1973), «The Reliability Coefficient for Maximum Likelihood Factor Analysis.» *Psychometrika* 38: 1-10.
98. Wang, L. L., X. Fan, and V. L. Wilson (1996), «Effects of Nonnormal Data on Parameter Estimates for a Model with Latent and Manifest Variables: An Empirical Study.» *Structural Equation Modeling* 3(3): 228-47.
99. Werts, C. E., and R. L. Linn (1970), «Path Analysis: Psychological Examples.» *Psychological Bulletin* 74: 193-212.
100. Wheaton, B. (1987), «Assessment of Fit in Overidentified Models with Latent Variables.» *Sociological Methods and Research* 16: 118-54.
101. Wheaton, B., D. Muthen, D. Alwin, and G. Summers (1977), «Assessing Reliability and Stability in Panel Models.» In D. Heise (ed.), *Sociological Methodology*, San Francisco: Jossey-Bass.
102. Williams, L. J., and J. T. Hazer (1986), «Antecedents and Consequences of Organizational Turnover: A Reanalysis Using a Structural Equations Model.» *Journal of Applied Psychology* 71 (May): 219-31.
103. Williams, L. J., and Patricia J. Holahan (1994), «Parsimony-Based Fit Indices for Multiple-Indicator Models.» *Structural Equation Modeling* 1(2): 161-89.
104. Wold, Herman, ed. (1981), *The Fixed Point Approach to Interdependent Systems*. Amsterdam: North Holland.
105. Wothke, W. (1993), «Nonpositive Definite Matrices in Structural Modeling.» In K. A. Bolden and J. S. Long (eds.), *Testing Structural Equation Models*, Newbury Park, Calif.: Sage.

La confianza del consumidor en el minorista: Desarrollo y contraste de un modelo conceptual

MANUEL SÁNCHEZ PÉREZ, M.ª ÁNGELES INIESTA Y ELVIRA SÁEZ GONZÁLEZ[1]
Universidad de Almería

Revista Española de Investigación de Marketing – ESIC, n.º 4

En el presente trabajo se analiza el concepto de confianza en las relaciones de intercambio, y se plantea un modelo conceptual de la confianza en el contexto de las relaciones minorista-consumidor. Se construyen diversas escalas para medir las diferentes dimensiones componentes del concepto confianza, contrastándose el modelo propuesto desarrollando a través de un análisis factorial confirmatorio de orden superior. La muestra de análisis asciende a 736 unidades, con recogida de datos mediante entrevista personal. Como conclusiones destacamos la importancia de la confianza para los minoristas y derivamos implicaciones estratégicas y tácticas para su gestión empresarial.

Palabras clave: *Confianza; relaciones comprador-vendedor; comercio minorista; análisis factorial confirmatorio; modelos estructurales.*

1. INTRODUCCIÓN

En las relaciones de intercambio, se generan pautas de comportamiento que ayudan a estabilizar las mismas y a reducir la incertidumbre. Se trata de vínculos normativos que permiten regular el comportamiento generando expectativas comunes para todas las partes (Pfeffer y Salancick, 1978). La confianza se configura como uno de estos vínculos clave que lleva a la adopción de comportamientos coordinativos (Clopton, 1984). La importancia de estos vínculos ha sido puesta de manifiesto por Macaulay (1963), quien estudiando las relaciones no contractuales entre empresas, encontró que estas normas, no sólo venían a rellenar los vacíos existentes en los contratos sino que muy a menudo operaban por encima de los términos escritos en un contrato. Asimismo, Williamson (1975) establece que los costes de intercambio pueden ser reducidos con la presencia de confianza en las transacciones.

El establecimiento de relaciones de marketing duraderas entre compradores y vendedores o entre las partes que intervienen en una alianza o asociación es una cuestión cada vez más presente tanto en la teoría como en la práctica de las organizaciones que está siendo analizada en los últimos tiempos bajo la filosofía del marketing de relaciones. Su estudio se ha centrado fundamentalmente en las áreas del marketing industrial y de servicios, si bien, se está extendiendo también al resto de campos de estudio del marketing. Concretamente, el estudio de las relaciones con los consumidores ha sido fundamentalmente desarrollado en el ámbito de los servicios (Grönroos, 1990). El intercambio relacional de una empresa con el consumidor final es una de las diez formas discretas de marketing de relaciones que distinguen Morgan y Hunt (1994). Estos autores posicionan la confianza y el compromiso como variables mediatizadoras clave entre los antecedentes y las consecuencias o resultados de una estrategia de marketing de relaciones, siendo su presencia básica para el éxito de la misma.

La confianza es la variable más aceptada universalmente como base para cualquier interacción o intercambio humano (Gundlach y Murphy, 1993). Ésta es una de las variables individuales de interés actual para los que investigan relaciones que puede verse, no como un rasgo estático de la misma, sino más bien como un esquema de la relación compartido por grandes segmentos de la población

[1] Los autores desean expresar su agradecimiento a los revisores anónimos por sus aportaciones que han permitido mejorar el trabajo.

(Berscheid, 1994). Webster (1992) considera la confianza y la credibilidad como los fundamentos básicos para el establecimiento de una relación. Para Grönroos (1989) el intercambio y cumplimiento de promesas y el mantenimiento y refuerzo de la confianza son dos elementos esenciales que deben estar presentes en la relación con el cliente.

Pese a la importancia de este concepto en las relaciones de intercambio, existe poca investigación empírica, siendo especialmente aguda esta carencia en las relaciones consumidor-minorista (último nivel del canal). Con el objetivo de elaborar un modelo conceptual de la confianza entre consumidor y minorista, en el presente artículo se hace una revisión de la investigación realizada sobre la confianza, fundamentalmente en las relaciones de intercambio. Se presentan distintas definiciones de confianza, dada la inexistencia de una definición universalmente aceptada en la ciencias sociales. Se procede después a analizar las características de esta variable y las distintas dimensiones que presenta, para desarrollar un modelo conceptual de la confianza entre consumidor y minorista, siendo a continuación contrastado con la construcción de un modelo factorial confirmatorio de orden superior. El trabajo finaliza con un conjunto de conclusiones.

2. LA CONFIANZA EN LA RELACIÓN DE INTERCAMBIO

La confianza es un concepto considerado desde distintas áreas de estudio. Esto ha dado lugar a que tanto su naturaleza como la definición y el ámbito en el cual se aplica, sean algo diferentes para cada disciplina. No obstante, la mayor parte de las aplicaciones se centran en el intercambio. De hecho, se ha dicho que la confianza es la variable más aceptada universalmente como base para cualquier interacción o intercambio humano (Gundlach y Murphy, 1993). Por ello, a pesar de la multidisciplinariedad del concepto no existen diferencias importantes entre unas consideraciones y otras. A continuación, desarrollamos brevemente las distintas perspectivas de estudio de la confianza.

La confianza ha sido ampliamente estudiada en el campo de la *psicología social* (Blau, 1964; Rotter, 1967; Deutsch, 1960). También ha sido tratada en el ámbito *económico* (Williamson, 1993), si bien, bajo el ideal de mercado neoclásico no era un concepto a tener en cuenta. Ha sido a raíz del énfasis del marketing en la construcción de relaciones más duraderas en el canal de comercialización, cuando se ha considerado la influencia de que un individuo o empresa confíe en que otro vaya a cumplir sus promesas. En la consideración de la confianza desde el campo de la *filosofía*, se enfatiza su carácter de actitud, definiéndose como una «actitud confiada», con frecuencia inconsciente, que es parte de la conducta básica de la vida (Blomqvist, 1997). En el ámbito *legal*, concretamente en materia de contratos de intercambio, se parte de la existencia de unas bases éticas en el intercambio, y los contratos incluyen aspectos como la solidaridad, la integridad en el rol, la mutualidad (Macneil, 1980), la equidad, la responsabilidad, el compromiso y la confianza (Gundlach y Murphy, 1993).

Por último, cabe destacar la presencia de la confianza en numerosas investigaciones de *marketing*, como una variable a tener en cuenta en el contexto del desarrollo del intercambio (Anderson y Narus, 1984; 1990; Dwyer, Schurr y Oh, 1987; Schurr y Ozanne, 1985).

Las distintas investigaciones llevadas a cabo en marketing en torno a este concepto tienen su origen en las diversas áreas de estudio mencionadas, si bien, la disciplina más influyente es la psicología social (Schurr y Ozanne, 1985). En ella se destaca la importancia de la confianza como un ingrediente esencial del acuerdo y la cooperación en el marco de las relaciones diádicas en particular y en el marco de la teoría del intercambio social en general. La psicología social parte de que la confianza lleva al diálogo constructivo y a la resolución de problemas de forma cooperativa.

Podemos citar las siguientes áreas en las que se ha tratado la confianza dentro del marketing:

a) En el marketing de servicios, la confianza ha sido tratada de forma extensa por diversos estudiosos del tema (Bendapudi y Berry, 1997; Crosby, Evans y Cowles, 1990; Moorman, Deshpandé y Zaltman, 1993; Moorman, Zaltman y Deshpandé, 1992; Parasuraman, Berry y Zeithaml, 1991).

b) También se encuentra presente la confianza en la teoría relativa a las relaciones en el canal de comercialización (Andaleeb, 1996; Anderson y Narus, 1984; 1990; Anderson y Weitz, 1989; Doney y Cannon, 1997; Ganesan, 1994; Morgan y Hunt, 1994; Schurr y Ozzane, 1985; Wilson, 1995).

c) En la literatura relativa a las estructuras alternativas de interrelación (acuerdos, alianzas...), la confianza aparece como un elemento necesario para la formación y el desarrollo de relaciones estables entre ellas (Bucklin y Sengupta, 1993; Madhok, 1995; Smith y Barclay, 1997; Webster, 1992).

d) La confianza está presente, tanto de forma implícita como explícita, en diversas teorías del liderazgo que discuten la modelización del rol, en las cuales, la confianza en el líder es una cuestión esencial a considerar (Rich, 1997).

e) En el ámbito del marketing industrial, la confianza también es considerada como un elemento central. Dentro del enfoque de interacción desarrollado por Håkansson y otros investigadores del IMP Group (Industrial Marketing and Purchasing Group), se estudian las relaciones entre los miembros del canal de comercialización, su naturaleza de largo plazo y la participación activa entre las partes, considerándose de forma implícita la confianza como uno de los distintos factores que inciden sobre ellas (Håkansson, 1982).

En la última década, también se ha desarrollado un cuerpo considerable de teoría alrededor del marketing de relaciones, donde se propone la confianza como un elemento básico, fundamental y necesario para que se establezcan y desarrollen relaciones a largo plazo y se pueda implementar un marketing de relaciones con éxito (Berry, 1995 Grönroos, 1994; 1996; Gurviez, 1997; Morgan y Hunt, 1994).

3. DEFINICIÓN DE CONFIANZA

La confianza ha sido estudiada, fundamentalmente, desde una perspectiva estática. Anderson y Narus (1990) establecen que cuando se pregunta sobre la confianza en una relación se suele contestar en base a lo que se siente en ese momento, sin considerar que el tiempo es sólo uno de los elementos a tener en cuenta cuando se usa la confianza como variable en la investigación de la relación. Autores como Ravald y Grönroos (1996) sí han considerado la dimensión temporal, entendiendo que la confianza es algo dinámico que, tras la realización de una serie de transacciones con éxito, se va desarrollando en el cliente.

En el contexto de las relaciones la confianza se establece mediante un escalamiento gradual y progresivo de las reciprocidades las cuales son adecuadamente concebidas y que se hacen operativas en unas relaciones de intercambio continuas (Shapiro 1987).

No existe una única definición de confianza que abarque todas las situaciones en las cuales aparece, ya sea la confianza en uno mismo, en alguien (individuo u organización), o en algo. Tampoco ha sido objeto de análisis la posibilidad de aplicar un único concepto para todos los niveles de confianza (personal, organizacional y cultural) ni tampoco la base ética de cada tipo de confianza (Murphy y Gundlach, 1997). En cualquier caso, la concepción de confianza considerada en este trabajo es la que se tiene hacia una persona o una organización, y se plantea a la hora de establecer relaciones, concretamente, relaciones de intercambio comerciales.

Murphy y Gundlach (1997) proponen los cinco tipos de confianza que exponemos a continuación, cada uno de los cuales puede ser dividido en tres niveles: Individual, organizacional y cultural.

1) *Confianza calculada*: Se basa en la economía. Ésta considera que la confianza se establece por medio de un proceso de cálculo en el que se computa el coste y/o beneficio que supondría el engaño de la otra parte o el de permanecer en la relación.
2) *Confianza verificable*: Se basa en la vulnerabilidad inherente asociada a la confianza. Así, la confianza implica un riesgo inherente de pérdida o vulnerabilidad.
3) *Confianza recíproca*: Enfatiza su naturaleza compartida. En los negocios existen relaciones de confianza, o una confianza mantenida por ambas partes de la relación.
4) *Confianza ganada*: Enfatiza las bases sobre las que se construye la confianza. Refleja el deseo de confiar en otra parte, adquirido a través de alguna base experiencial (percepciones, experiencias pasadas, reputación, conocimiento, experiencia o valores ideológicos).
5) *Confianza ciega o incuestionable*: Es, posiblemente, una forma de confianza irracional, en la que se ejerce una lealtad permanente y ciega. Consiste en confiar en situaciones que mucha gente pensaría que no garantizan confianza.

Las diferentes definiciones de la confianza pueden ser agrupadas sobre la base de sus características comunes. En un primer grupo, y partiendo de la literatura de la Psicología Social, Rotter (1967:651) la definió como «la esperanza generalizada sostenida por un individuo de que se puede fiar de la palabra del otro». Posteriormente, Schurr y Ozanne (1985: 940) han definido la *confianza* como «la creencia de que la palabra o promesa de una parte es firme y segura y que esa parte cumplirá sus obligaciones en una relación de intercambio». Por otro lado, Gambetta (1988: 217) define la confianza como «un nivel particular de probabilidad subjetiva con el cual un agente valora que otro agente o grupo de agentes llevarán a cabo una acción particular. Cuando decimos que confiamos en alguien o que alguien es digno de confianza, implícitamente queremos decir que la probabilidad de que esa persona lleve a cabo una acción beneficiosa o, al menos, no negativa para nosotros, es bastante alta como para que consideremos unirnos en alguna forma de cooperación con él». Estas tres definiciones coinciden en considerar la confianza como una expectativa de comportamiento futuro.

En un segundo grupo, englobamos una serie de definiciones surgidas dentro del ámbito de las relaciones interorganizacionales, como las de Anderson y Narus (1990) y Andaleeb (1996), las cuales coinciden en la consideración de la confianza como la creencia en la obtención de resultados positivos. Las situaciones de confianza en gran parte de la literatura relevante son entendidas como una subclase de aquéllas que implican riesgo, es decir, aquéllas en las que el riesgo de una parte depende de la actuación de la otra (confianza verificable). De este modo, la confianza está garantizada cuando la ganancia esperada de situarse en una posición de riesgo respecto a otra, es positiva, pero no en otro caso (Williamson, 1993). En este segundo grupo, la orientación 'confianza calculada' es evidente.

En un tercer grupo, agrupamos un conjunto de definiciones cuya característica común es que en todas ellas está presente algún modo de confianza recíproca, al hacer todas ellas referencia a las relaciones entre dos o más partes en un intercambio o conjunto de intercambios (Anderson y Weitz, 1989; Moorman *et al.*, 1992; Moorman *et al.*, 1993; Morgan y Hunt, 1994; Smith y Barclay, 1997).

Considerando estas diferentes perspectivas, una definición de confianza integradora y más completa que englobe todas las pautas en las que vienen a coincidir las diferentes definiciones debe considerar que la confianza es:

- La creencia o expectativa
- de que se produzcan determinado tipo de efectos: cumplimiento de obligaciones, comportamiento esperado, resultados positivos, satisfacción de necesidades
- para/por una o todas las partes que participan en una relación social o de intercambio (ya sea a lo largo del canal de comercialización o en los acuerdos entre organizaciones)
- y que desemboca en la intención de desarrollar dicha relación.

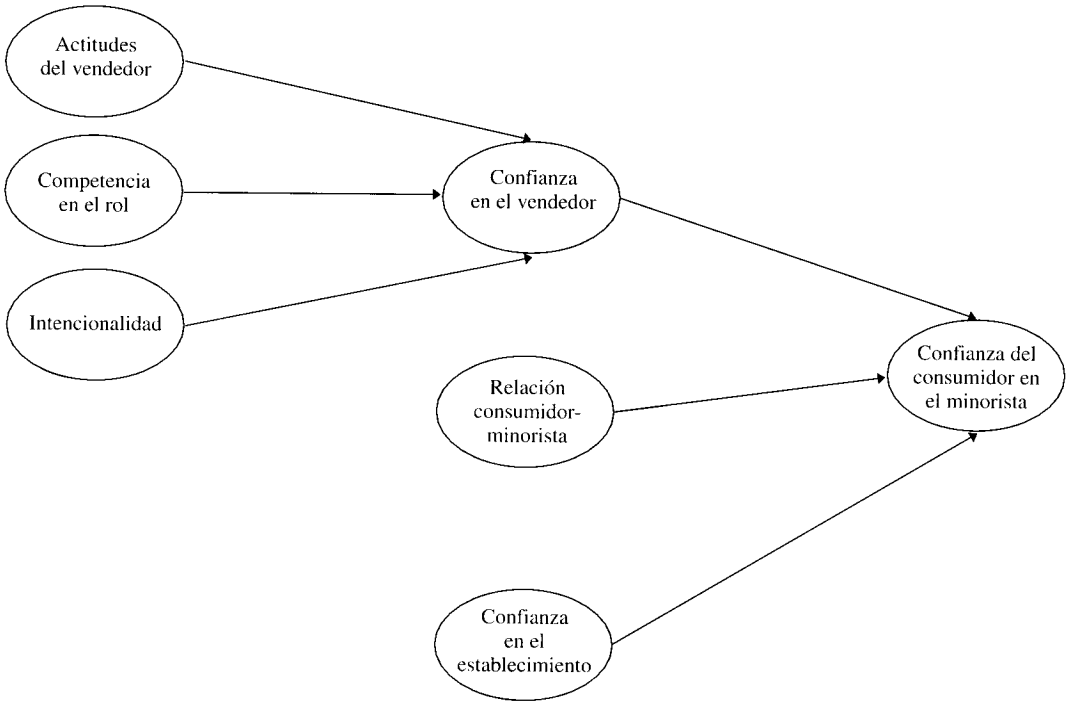


FIGURA 11.L.1. Modelo propuesto de confianza entre consumidor y minorista

4. UN MODELO DE CONFIANZA ENTRE CONSUMIDOR Y ESTABLECIMIENTO MINORISTA

Los estudios analizados en el apartado anterior consideran la confianza en diversos contextos, pero todos ellos referidos a diferentes tipos de relaciones entre empresas del canal de comercialización, tanto a nivel horizontal como vertical. Del análisis de la literatura relativa a la confianza en el contexto de las relaciones de intercambio, se desprende la inexistencia de investigaciones empíricas aplicadas a las relaciones entre los minoristas y los consumidores finales.

Se propone un modelo en el que, partiendo de la investigación desarrollada en marketing (aplicada fundamentalmente a las relaciones de intercambio entre organizaciones), se analizan y justifican las diversas dimensiones que configuran la confianza del consumidor en el minorista como construcción de marketing. En el modelo se considera que la confianza está compuesta por tres dimensiones básicas: 1) la confianza percibida del consumidor en el vendedor, 2) la relación consumidor-minorista y 3) la confianza percibida del consumidor en el establecimiento, dentro de las cuales se engloban un conjunto de factores que configuran el concepto de confianza (Figura 11.L.1).

4.1. Confianza percibida del consumidor en el vendedor

Dentro de esta dimensión global distinguimos tres grupos de variables con incidencia en la confianza: actitudes del vendedor, competencia en el rol e intencionalidad.

4.1.1. Actitudes del vendedor

Dentro de este grupo, la **formalidad** es propuesta como un antecedente de la confianza por Moorman *et al.* (1993), quienes establecen que ésta última aumenta con la formalidad del vendedor cuando los clientes cuentan con que las acciones de aquél son previsibles y consistentes. El análisis empírico de estos mismos autores no confirmó la existencia de relación entre ambas variables. No obstante, ha sido incluida en nuestro análisis, pues el cambio de contexto de la investigación puede llevar a una relación positiva entre ambas.

La **responsabilidad** es considerada por Smith y Barclay (1997) como un atributo de la 'confiabilidad percibida mutua', dentro de la dimensión que ellos llaman 'carácter'. Por tanto, consideramos que en la medida en que el cliente perciba que el vendedor es responsable, eliminará su temor a un comportamiento oportunista y le llevará a depositar su confianza en él.

La **confiabilidad** es considerada por Smith y Barclay (1997) como una de las facetas de la confianza, junto con el 'comportamiento de confianza' o deseo de unirse a tal comportamiento, debiendo estar ambos presentes y en ambas partes de la relación de intercambio si se pretenden conseguir relaciones continuas y estables (Anderson y Weitz, 1989). Siguiendo a Moorman *et al.* (1993), la confianza como una creencia o expectativa sobre la confiabilidad de la otra parte del intercambio es una de las aproximaciones de la confianza, la cual resulta de la experiencia, formalidad o intencionalidad de la otra parte. Basándonos en esos tres

atributos, hemos incluido en el modelo la expectativa sobre la confiabilidad que tienen los consumidores como una variable que repercute positivamente sobre la confianza en la relación comprador-vendedor.

La **integridad** del vendedor, entendida como el deseo de no sacrificar los estándares éticos para lograr objetivos personales, también ha sido considerada como un factor relacionado con la confianza por Moorman *et al.*(1993), resultando de su estudio una relación positiva entre ambas. En nuestra investigación, un vendedor que demuestre integridad es probable que inspire confianza en el comprador, ya que éste puede esperar que aquél no renuncie a los estándares de la ética profesional y personal.

La **afinidad**, entendida como la valoración del comprador de que el vendedor es amigable, amable y agradable, afecta positivamente a la confianza (Doney y Cannon, 1997), ya que un comprador confiará más en sus predicciones sobre alguien a quien ve afín o por quien siente simpatía.

La **credibilidad** supone una esperanza de que se puede creer en el mensaje emitido por la otra parte del intercambio. Ganesan (1994), encuentra que la credibilidad se ve influida por la reputación y la percepción del comprador de las inversiones específicas que el vendedor realiza en la transacción, aumentando la confianza global. Planteamos, por tanto, una relación positiva entre credibilidad y confianza.

4.1.2. Competencia en el rol

Según Smith y Barclay (1997: 6), la competencia en el rol es «el grado en que las partes perciben que el otro tiene las destrezas, las habilidades y el conocimiento necesario para desempeñar su labor de forma efectiva». Este grupo engloba siete variables, que pasan a exponerse a continuación.

Los **conocimientos técnicos** abarcan el propio conocimiento y la competencia técnica del vendedor, siendo un factor clave que afecta la confianza (Moorman *et al.*, 1993; Bendapudi y Berry, 1997). Crosby *et al.* (1990) establecen una relación positiva entre la experiencia percibida y la calidad de la relación. Por un lado, la confianza es una de las dimensiones de la calidad de la relación que ellos proponen. Por otro lado, establecen que la percepción de la experiencia del vendedor refleja la identificación de competencias relevantes asociadas con los bienes o servicios, mostrada frecuentemente en forma de información proporcionada por el vendedor. Por tanto, podemos plantear una relación positiva entre los conocimientos técnicos que posee el vendedor y la confianza del comprador.

Otra variable considerada ha sido la **habilidad** del vendedor percibida por el consumidor, la cual aumenta su confianza en aquél, pues le mueve a fiarse más de la palabra el vendedor en cuanto al cumplimiento de sus promesas (Crosby *et al.*, 1990; Moorman *et al.*, 1993; Doney y Cannon, 1997).

Otra característica relacionada con un estilo de venta orientada a la persona (Kotler, Cámara y Grande, 1995) ha sido su **accesibilidad**, que hace referencia a la facilidad de llegar o acercarse a alguien que tiene un trato fácil. Como tal, este atributo no ha sido encontrado en la literatura que analiza la confianza en marketing, si bien, se ha considerado entendiendo que, en el ámbito de las relaciones entre un comprador final y el vendedor del establecimiento, esta cualidad del minorista podría favorecer la formación de confianza en la relación.

La **disponibilidad** o predisposición a responder a las necesidades de los usuarios en sus requerimientos de información, es decir, estar a su disposición, es una característica propia de la relación con el vendedor que no se ha considerado en el contexto de las relaciones interorganizacionales, pues tal y como se ha definido, no tiene mucho sentido considerarla como un factor independiente en dicho escenario. Sin embargo, dentro del ámbito en que nos movemos en este trabajo, éste sí es un factor que puede afectar a la confianza del consumidor en el vendedor de un establecimiento. Éste concepto deriva de factores anteriormente expuestos como la habilidad, afinidad/simpatía o responsabilidad.

La **autoridad en el establecimiento** por parte del vendedor, o la creencia por parte del consumidor de que aquél puede proporcionarle lo que le dice o promete porque tiene autoridad para hacerlo sobre los recursos de la empresa, lleva al comprador a fiarse más de la palabra del vendedor, afectando positivamente a la confianza (Moorman *et al.*, 1993; Doney y Cannon, 1997). Este poder viene relacionado con el nivel de experiencia, de información y de tecnología existente (Doney y Cannon, 1997).

El **trato individualizado** lleva al vendedor a dirigirse de forma particular y personalizada a cada consumidor influyendo sobre la confianza de los clientes. En su consecución, las bases de datos de marketing juegan un papel esencial (Cespedes y Smith, 1993). Doney y Cannon (1997) analizan la repercusión de esta variable en la confianza, si bien, al aplicarlo en las relaciones del canal de comercialización la consideran en el sentido de realizar inversiones específicas en la empresa compradora. De su estudio empírico, esta variable no resultó significativa a pesar de lo cual, nosotros hemos considerado el trato personalizado hacia el cliente, hipotetizando un efecto positivo del mismo sobre la confianza.

La ***oportunidad de respuesta***, o la percepción del consumidor respecto a la eficiencia del vendedor para responder a sus necesidades en el momento adecuado, es un factor relacionado con la percepción de la calidad de servicio que puede aumentar su confianza (Moorman *et al.*, 1993).

4.1.3. Intencionalidad

Dentro de este grupo se engloban, por un lado, la creencia por parte del comprador de que el vendedor es capaz de decidir y actuar de forma que se fomenten los intereses conjuntos, y por otro, la medida en que se perciben los propósitos que hay detrás de las acciones del vendedor. La intencionalidad viene caracterizada por los siguientes factores.

En primer lugar y como característica del vendedor se ha considerado la **sinceridad**, o percepción sobre su honestidad e intención en el cumplimiento de promesas, habiéndose de-

tectado que afecta positivamente sobre la confianza (Moorman *et al.*, 1993). El mantenimiento de promesas constituye un medio importante de lograr la satisfacción del cliente, su retención y una rentabilidad a largo plazo (Grönroos, 1994), lo cual debe repercutir positivamente sobre la confianza. Por tanto, planteamos que se debe encontrar una relación positiva entre la sinceridad y la confianza.

La **semejanza** repercute de forma positiva sobre la confianza (Doney y Cannon, 1997). Engloba la creencia por el comprador de que el vendedor comparte intereses comunes y valores con él, lo cual puede hacer que el comprador se sienta más capaz de valorar las intenciones del vendedor y de predecir su comportamiento futuro. Morgan y Hunt (1994) señalan que los valores comunes son un precursor directo de la confianza. El estudio realizado por Dwyer *et al.* (1987) sugiere que la semejanza actitudinal y de valores fomenta la confianza hacia los demás. También, Bendapudi y Berry (1997) sostienen que la semejanza con otras partes las hace más atractivas, gustan más y son más dignas de confianza que otras con las que no se comparten valores comunes, proponiendo una relación positiva entre ambas variables. Por su parte, Crosby *et al.* (1990) establecen una relación positiva entre semejanza y calidad de la relación, dentro de la cual la confianza es una de las dimensiones consideradas.

En la literatura se reconoce con unanimidad la importancia y relación de la confianza en la **cooperación**. Pero existe divergencia acerca de su consideración como causa o efecto. Dwyer *et al.* (1987) consideran la confianza como el concepto clave en la comprensión de las expectativas de cooperación. Anderson y Weitz (1989) establecen que el apoyo que un miembro recibe del otro, es decir, la cooperación, permite la existencia de confianza. Anderson y Narus (1990), si bien proponen inicialmente que la confianza es causa de la cooperación, el desarrollo empírico posterior les llevó a considerar la cooperación como causa antecedente de la confianza, y en este último sentido la recogemos en este estudio.

La **orientación colectiva** recoge el deseo del vendedor de cooperar con el comprador, facilitándole información y consejo acerca de lo que más le puede interesar. Este factor fue considerado por Moorman *et al.* (1993), quienes propusieron una relación positiva con la confianza; sin embargo, su estudio no dió ninguna relación significativa entre ellos.

El **oportunismo** (Williamson, 1975), o grado en que no se actúa con espíritu de cooperación, engañando y negando una acción útil, es una práctica cuyo desarrollo por parte del vendedor repercute de forma negativa sobre la formación de confianza (John, 1984; Smith y Barclay, 1997).

4.2. Relación consumidor-minorista

Como acabamos de exponer, en el proceso de generación de confianza hacia el minorista, el vendedor representa una fuente importante de ésta para el consumidor, al igual que las políticas, acciones y oferta del establecimiento. Este concepto, propuesto para el ámbito de los servicios, implica que cualquier interacción entre un cliente y el proveedor de un servicio tiene el potencial para fortalecer, debilitar o, incluso, destruir la relación entre ellos (Bendapudi y Berry, 1997). Aplicado al ámbito del comercio minorista, la configuración de la relación cliente-minorista, es otro camino que lleva a la construcción de relaciones duraderas, debiendo ser la creación de confianza uno de sus primeros objetivos. Considerando el ámbito de aplicación de este estudio, hemos considerado los siguientes factores. La **longitud de la relación** presenta una relación positiva con la confianza al aumentar la habilidad del comprador para predecir el comportamiento futuro del vendedor (Anderson y Weitz, 1989; Doney y Cannon, 1997).

El **contacto de negocios o social con el vendedor** facilita al consumidor la observación de su comportamiento, lo cual le da más oportunidades para evaluarlo. A mayor frecuencia, más serán las situaciones observadas, y en la medida en que dichos contactos sean satisfactorios la confianza aumentará (Doney y Cannon, 1997; Bendapudi y Berry, 1997). Por su parte, Crosby *et al.* (1990) también encontraron una relación positiva entre la frecuencia de los contactos y la confianza en el vendedor. Estas hipótesis no dieron los resultados esperados, verificándose sólo la relación positiva entre el contacto de negocios y la confianza. Los razonamientos de Doney y Cannon (1997) y Bendapudi y Berry (1997) que se acaban de exponer se pueden hacer extensibles a la variable **frecuencia de compra** y, por tanto, a mayor frecuencia de compra, mayores serán las situaciones observadas, pudiendo conducir a una mayor confianza en la medida en que dichos contactos sean satisfactorios.

Los **lazos sociales** entre comprador y vendedor, ya sea dentro de la relación de intercambio (por ejemplo, proporcionando tarjetas o regalos) o fuera de la relación de negocio (como los encuentros en un club al que ambos pertenecen o realizando obras sociales en las que las dos partes intervienen), pueden influir sobre la confianza del cliente en el vendedor (Bendapudi y Berry, 1997), reduciendo o eliminando el riesgo de un comportamiento oportunista.

En la literatura del comportamiento del consumidor se reconoce la influencia que tienen la familia, amigos y grupos de referencia en el comportamiento de compra (Childers y Rao, 1992), pudiendo influir sobre la confianza del consumidor en el vendedor o el establecimiento. Por tanto, la **experiencia de otros** se propone como un factor relacionado positivamente con la confianza (Schurr y Ozanne, 1985; Bendapudi y Berry, 1997).

También en los estudios del comportamiento del consumidor, se ha considerado la **experiencia directa de compra** como una variable clave en los procesos de toma de decisiones y de elección por parte del consumidor (Alba y Hutchinson, 1987), que repercute positivamente en el desarrollo de confianza por parte de éste hacia los productos, el vendedor y el establecimiento (Schurr y Ozanne, 1985).

4.3. Confianza percibida por el consumidor en el establecimiento

En el modelo se han considerado como indicativas de una mayor confianza once acciones distintas del establecimien-

to minorista. Las **estrategias de influencia**, o el uso de las bases de poder por parte del establecimiento, varían en función de la competencia o factores que caracterizan la estructura relacional (Frazier, 1983). Dichas estrategias de influencia, cuando no son coercitivas, pueden afectar significativamente a la confianza. Partiendo de esto, vamos a considerar el ejercicio de estrategias de influencia no coercitivas por parte del minorista como un factor relacionado positivamente con la confianza del cliente.

La **inversión en la relación** hace referencia a los recursos o el esfuerzo dedicados a la relación, sin valor fuera de ésta, y que no pueden recobrarse si se acaba. La parte inversora, por tanto, pierde cuando se disuelve la relación, porque las inversiones no tienen valor de mercado. La inversión en la relación ha sido considerada como una variable indicativa de confianza, con la que se relaciona positivamente (Smith y Barclay, 1997; Bendapudi y Berry, 1997).

La **apertura de la comunicación** se refiere al grado en que se comparte información oportuna de manera formal o informal. Cuando las comunicaciones pasadas de una fuente son dignas de confianza, la probabilidad de que los receptores de la información se fíen más de las comunicaciones actuales de dicha fuente es mayor (Moorman *et al.*, 1993). En la investigación realizada por Smith y Barclay (1997), esta variable resultó ser un predictor importante del comportamiento de confianza. Por su parte, Berry (1995) establece que una comunicación regular, abierta y bilateral, conduce los objetivos de la empresa hacia la consecución del bienestar de los clientes, generando confianza.

El **oportunismo** se ha considerado también como una variable relacionada negativamente con la confianza en el contexto de la relación con el establecimiento, al poder ser percibida dicha conducta en sus políticas generales, no sólo en los vendedores.

La **política de resolución de conflictos** (Dant y Schul, 1992) puede tener un efecto positivo sobre la confianza en el minorista. El conflicto en un intercambio, ya sea éste relacional o discreto, tendrá distinta repercusión sobre la confianza en función de cuáles sean sus causas, el tipo de conflicto desarrollado y su intensidad (Stern y El-Ansary, 1992). Schurr y Ozanne (1985) sostienen que mayores niveles de confianza producen mayores niveles de comunicación, una actitud más favorable hacia la resolución de problemas e incrementa el compromiso de las partes hacia la implementación de acuerdos. Por consiguiente, planteamos en el modelo una relación positiva entre la existencia de una adecuada política de resolución de conflictos y la confianza.

La adaptación es un rasgo característico en la dinámica de las relaciones de negocio. En nuestro contexto se hace referencia al grado en que el minorista se acomoda a las necesidades del cliente, distinguiendo, por un lado, la **aceptación de influencia** y, por otro, la **aceptación de sugerencias**. En ambos casos aplicamos la literatura de las estrategias de influencia pero, en este caso, es el consumidor el que las ejerce. El minorista puede estar dispuesto a aceptar la influencia del consumidor y adaptarse a sus requerimientos en cuestión de horarios, ofertas, etc., así como aceptar sus sugerencias en aspectos como la ubicación de los productos, la decoración del establecimiento, etc. Una gestión de negocios eficiente implica desarrollar una cierta orientación hacia el mercado, estando dispuesta la empresa a adaptarse al consumidor (Kohli y Jaworski, 1990). Las partes pueden hacer adaptaciones para conseguir un ajuste inicial entre sus necesidades y capacidades, pero éstas también son necesarias en una relación continua (Hallén, Johanson y Seyed-Mohamed, 1991). Teniendo en cuenta estas consideraciones, en el modelo se plantean tanto la aceptación de influencia como la aceptación de recomendaciones como dos factores que repercuten positivamente sobre la confianza.

La **notoriedad** del nombre comercial del establecimiento desarrollada a través de sus actividades de comunicación comercial puede influir positivamente sobre la confianza ya que desde el hecho de realizar publicidad en medios convencionales, hasta la publicidad en el lugar de venta, pasando por las promociones en el punto de venta y el buzoneo que habitualmente realizan, contribuyen a fortalecer la imagen del establecimiento y a crear seguridad en el mismo en la mente de los consumidores, seguridad que suponemos puede incidir positivamente sobre la confianza.

Otro factor a considerar es el **nivel de servicios**. Cada vez más los consumidores van buscando no sólo bienes y servicios sino una oferta holística de servicio que incluye todo lo relacionado con la entrega, instalación, actualización, reparación, mantenimiento o cómo usar mejor y de forma más segura el producto físico. Pero, además, demandan que les sea ofrecido de una forma más amigable, oportuna y confiable (Grönroos, 1996). Schurr y Ozanne (1985) comentan varias prácticas de marketing que implícitamente estimulan el que el consumidor confíe en el vendedor, entre las que se encuentran las marcas, logotipos y garantías, los cuales implican confianza. La prestación adecuada de los servicios ofrecidos puede simbolizar el compromiso de la empresa a desempeñar su papel de forma adecuada con los clientes, facilitando la diferenciación competitiva e incrementando la confianza del consumidor (Berry, 1995).

En lo relativo a la **calidad**, ésta es un factor multidimensional que está configurado por varios atributos como la responsabilidad, conocimientos técnicos, autoridad, personalización o comunicación (Parasuraman, Zeithaml y Berry, 1985), aspectos, todos ellos, contemplados en nuestro modelo en las dimensiones que configuran la confianza en el minorista. Por otra parte, en un estudio realizado en EEUU sobre los consumidores, se detectó que los productos de alta calidad poseen como primera característica percibida por el comprador que son dignos de confianza. Aquí puede verse, también, la relación existente entre ambos conceptos. No obstante, y dependiendo de cuál sea el primer factor que desarrolle el consumidor, se puede proponer la relación inversa, partiendo de que éste conoce la calidad de los productos del establecimiento y ésta le ayuda a desarrollar una mayor confianza en él.

La **calidad de servicio** puede ser un factor que genere confianza en el consumidor hacia el establecimiento que ofrece dichos servicios. Ciertamente, para que la prestación de servicios adicionales que acompañan la venta de bienes constituya un factor diferenciador de la empresa, éstos deben ser de calidad, y dicha calidad debe ser percibida por el consumidor para conseguir ese objetivo de diferenciación.

TABLA 11.L.1. Ficha técnica

Ámbito	Capital de Almería
Universo	Censo de comercios minoristas con códigos IAE comprendidos entre 641 y el 665 (2906 unidades)
Muestra	736 entrevistas personales con un error de muestreo del 3,68 % para un nivel de confianza del 95,45% (K=2 y p=q=0,5)
Selección	Muestreo aleatorio por código IAE (del 641 al 665) estratificado con afijación proporcional al número de establecimientos de cada tipo
Informante	Un cliente, seleccionado aleatoriamente, de cada comercio seleccionado
Trabajo de campo	Febrero de 1998
Entrevista	Personal mediante cuestionario estructurado

Cada vez que el consumidor interactúa con la empresa se produce un encuentro que puede contribuir a incrementar la satisfacción y el deseo del consumidor de continuar relacionándose con ella. La empresa puede aprovechar cada encuentro para proporcionar servicios de calidad, creando así confianza e incrementando la lealtad del cliente, y favoreciendo de este modo la construcción de una relación sólida (Bitner, 1995). Dado que los vendedores y demás empleados de un establecimiento son los que implementan el servicio al cliente, el buen desempeño de su labor constituye un factor clave de éxito en una empresa orientada al consumidor, especialmente en la prestación de los servicios que acompañan la venta de bienes (Beatty, Mayer, Coleman, Reynolds y Lee, 1996). Por tanto, en nuestro modelo hemos propuesto que la calidad de servicio y la confianza se encuentran relacionadas, siendo dicha relación positiva.

5. METODOLOGÍA

5.1. Muestra y datos

Para la contraste empírica del modelo de confianza consumidor-minorista propuesto se diseñó una muestra de 736 establecimientos minoristas pertenecientes a Almería capital y seleccionados de forma aleatoria con afijación proporcional por código IAE según tamaño de cada tipo (Tabla 11.L.1).

Se realizó una entrevista personal a un cliente de cada uno de los comercios aleatoriamente seleccionados. Después de realizar un pretest, el cuestionario utilizado estaba compuesto de 45 cuestiones relativas a las seis subescalas consideradas en el modelo de la Figura 11.L.1. Para estas preguntas se utilizó una escala de cinco niveles de categorías de respuesta equilibradas («Muy en desacuerdo» a «Muy de acuerdo»). Asimismo, se incluían diversos datos de identificación de los respondientes (sexo, edad, domicilio, teléfono) y del minorista (tamaño, sucursales, antigüedad).

5.2. Evaluación de las medidas

Cada una de las medidas fue sometida a un análisis de fiabilidad y validez siguiendo el procedimiento secuencial propuesto por Churchill (1979).

Una vez generada la muestra inicial de ítemes y recogidos los datos se procedió a depurar las medidas propuestas en el modelo teórico. Para ello, los ítemes fueron analizados combinando análisis factorial y correlaciones item-puntuación total de la escala. Aquellos ítemes que mostraban bajas cargas factoriales o con correlaciones item-total de la escala inferiores a 0,20 fueron eliminados[2]. En la Tabla 11.L.2 se muestran los tamaños de las medidas resultantes y las correlaciones item-total media para cada medida. Las dimensiones consideradas y las variables que las componen se recogen en la Tabla 11.L.3.

No obstante, las medidas fueron sometidas a mayor análisis con objeto de confirmar la composición resultante tras el proceso de depuración. Cada una de las subescalas fue sometida a un análisis factorial confirmatorio utilizando EQS 5.1. (Bentler, 1995) y LISREL 8 (Jöreskog y Sörbom, 1993). Como método de estimación se utilizó máxima verosimilitud sobre la matriz de covarianzas, obteniendose resultados similares con ambos softwares. El ajuste de cada uno de los modelos fue evaluado mediante siete índices: average off-diagonal absolute standardized residuals (AOASR), Bentler-Bonett normed fit index (BBNFI), comparative fit index (CFI), goodness-of-fit index (GFI), adjusted-goodness-of-fit index (AGFI), incremental fit index (IFI) y normed fit index (NFI)[3]. Además, se verificó la significatividad de las cargas. Los resultados obtenidos son satisfactorios para todas las subescalas de acuerdo con los valores generalmente reco-

[2] Este procedimiento de depuración de la escala es el utilizado por John y Reve (1982).

[3] Algunos de estos índices fueron estimados mediante EQS mientras otros fueron determinados con LISREL 8.

TABLA 11.L.2. Composición de las subescalas tras el proceso de depuración

Subescala	*Número de ítemes*	*Correlación Ítem-total media*
Actitudes del vendedor (CRV)	6	0,6323
Competencia en el rol (CER)	7	0,407
Intencionalidad (INT)	4	0,348
Relación consumidor-minorista (REL)	7	0,721
Confianza en el establecimiento (CRE)	11	0,339

TABLA 11.L.3. Dimensiones y medidas del modelo teórico propuesto sobre la confianza

ACTITUDES DEL VENDEDOR	Formalidad (X_1) Integridad (X_2) Responsabilidad (X_3) Confiabilidad (X_4) Afinidad (X_9) Credibilidad (X_{12})
COMPETENCIA EN EL ROL	Conocimientos técnicos (X_7) Habilidad (X_8) Accesibilidad (X_{10}) Disponibilidad (X_{11}) Autoridad en el establecimiento (X_{13}) Trato individualizado (X_{16}) Oportunidad de respuesta (X_{20})
INTENCIONALIDAD	Sinceridad (X_5) Semejanza (X_6) Orientación colectiva (X_{21}) Oportunismo (X_{22})
RELACIÓN CONSUMIDOR-MINORISTA	Longitud de la relación (X_{27}) Contacto de negocios (X_{28}) Contacto social (X_{29}) Frecuencia de compra (X_{30}) Lazos sociales (X_{31}) Experiencia de otros (X_{14}) Experiencia directa de compra (X_{15})
CONFIANZA EN EL ESTABLECIMIENTO	Estrategias de influencia (X_{17}) Inversión en la relación (X_{18}) Apertura de comunicación (X_9) Oportunismo (X_{23}) Política de resolución de conflictos (X_{24}) Aceptación de influencia (X_{25}) Aceptación de sugerencias (X_{26}) Notoriedad (X_{32}) Adecuación del nivel de servicios (X_{33}) Importancia de la calidad (X_{34}) Importancia de la calidad de servicio (X_{35})

mendados (Tabla 11.L.4). La composición de las medidas se consideró, pues, aceptable, quedando confirmada la composición de las subescalas.

Además de evaluar la consistencia interna con el coeficiente alpha, se determinó la fiabilidad compuesta (Fornell y Larcker, 1981) y el índice de Werts, Rock, Linn y Jöreskog (1978) (Tabla 11.L.5). Para calcular este último se utilizaron las matrices de covarianzas estimadas ($\sum$) y de varianzas residuales obtenidas por LISREL 8. Los coeficientes alpha son aceptables en general. Los valores meno-

TABLA 11.L.4. Índices de evaluación del análisis factorial confirmatorio sobre cada subescala

Subescala	*AOASR*	*BBNFI*	*CFI*	*GFI*	*AGFI*	*IFI*	*NFI*	*Significatividad de las cargas*
CRV	0,028	0,971	0,977	0,98	0,95	0,97	0,97	$p < 0,01$
CER	0,054	0,860	0,873	0,95	0,89	0,86	0,84	$p < 0,01$
INT	0,090	0,772	0,773	0,96	0,78	0,80	0,79	$p < 0,01$
REL	0,030	0,958	0,964	0,96	0,93	0,96	0,95	$p < 0,01$
CRE	0,044	0,820	0,848	0,95	0,92	0,83	0,80	$p < 0,01$

TABLA 11.L.5. Indicadores de fiabilidad: alpha, fiabilidad compuesta e índice de Werst *et al.*

Subescala	α	*Fiabilidad compuesta*	*Índice de Werts* et al.
CRV	0,847	0,850	0,937
CER	0,686	0,704	0,689
INT	0,558	0,586	0,695
REL	0,832	0,843	0,847
CRE	0,70	0,686	0,662

TABLA 11.L.6. Correlaciones entre las seis subescalas observadas

	CRV	*CDR*	*INT*	*REL*	*CRE*
CRV	0,847				
CER	0,673	0,686			
INT	0,611	0,525	0,558		
REL	0,458	0,350	0,452	0,832	
CRE	0,567	0,509	0,477	0,568	0,70

Todas las correlaciones son significativamente distintas de cero para un nivel del 5%
Coeficiente α en la diagonal principal

res para INT son aceptables considerando la longitud de las escalas y el carácter del estudio (Nunnally, 1987; Peterson, 1994). El nivel en torno a 0,7 recomendado para la fiabilidad compuesta se alcanza en la mayor parte de los casos. Por las razones antes mencionadas es algo menor también para INT. Los resultados del índice de Werts *et al.*, son, asimismo, aceptables. Finalmente, todas las cargas de las seis escalas resultaron significativas ($p < 0,01$).

La validez de contenido ha quedado verificada por la composición de las escalas realizadas a través de una revisión bibliográfica exhaustiva, habiéndose utilizado las variables observadas en diversas investigaciones anteriores relacionadas con nuestro objeto de investigación.

Una vez depuradas las subescalas y contrastada la fiabilidad de las medidas observadas, procederemos a confirmar la existencia de validez del concepto confianza mediante el análisis de la validez convergente y la validez discriminante. La validez discriminante es evaluada comprobando que las correlaciones entre las subescalas no sean tan altas como su coeficiente alpha. Como puede verse en la Tabla 11.L.6, la validez discriminante está perfectamente confirmada.

En esencia, la validez convergente de las medidas es puesta de manifiesto por la significatividad de todas las cargas factoriales, como se recoge en la Tabla 11.L.4 (John y Reve, 1982). La Tabla 11.L.7 muestra la composición de la varianza de cada subescala entre varianza verdadera y varianza del error. La varianza verdadera es razonablemente alta excepto para CER, INT y CRE. En suma, los datos parecen mostrar validez convergente.

TABLA 11.L.7. Composición de la varianza de la escala

Subescala	*Varianza Verdadera*	*Varianza del error*
CRV	0,49	0,51
CER	0,28	0,72
INT	0,27	0,73
REL	0,46	0,36
CRE	0,21	0,79

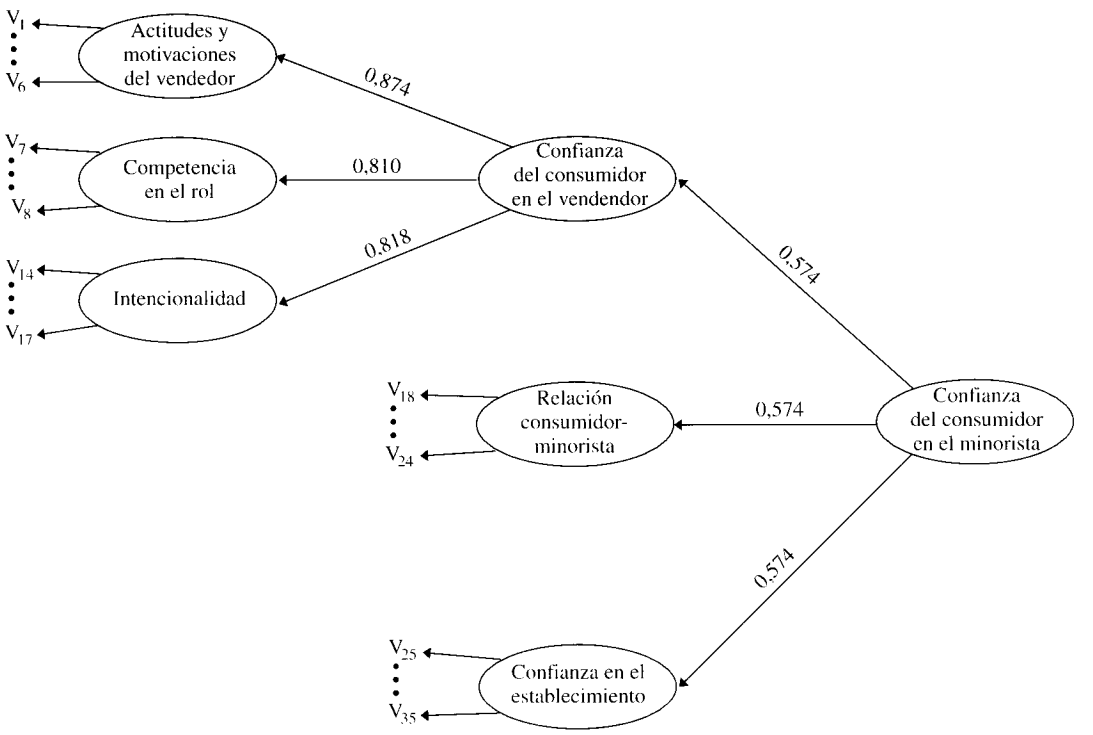


FIGURA 11.L.2. Resultados de las componentes de la confianza consumidor-minorista: relaciones factoriales de segundo y tercer orden establecidas en el modelo (no estandarizados)

5.3. Análisis del modelo factorial de orden tres especificado

El modelo de confianza percibida del consumidor hacia el minorista constituye un modelo factorial de orden tres. En un primer nivel, a partir de cada una de las subescalas construidas se extraen los correspondiente factores de orden uno: F_1 (actitudes y motivaciones del vendedor), F_2 (competencia en el rol), F_3 (intencionalidad), F_4 (relación consumidor-minorista), y F_5 (confianza en el establecimiento). A su vez, y en un segundo nivel, los tres primeros se corresponden con el factor de orden dos F_6 (confianza del consumidor en el vendedor) que ya en un tercer nivel, junto con F_4 y F_5 (éstos en segundo nivel) se encuentran relacionados bajo F_7 (confianza del consumidor en el minorista). La Figura 11.L.2 ilustra el modelo propuesto.

La estimación del modelo propuesto se llevó a cabo mediante EQS 5.1., utilizando el método de máxima verosimilitud. Este programa permite estimar modelos de estructuras de covarianzas a partir del concepto amplio Bentler-Weeks de variable dependiente e independiente. Constituye, pues, un enfoque unificado de análisis de covarianzas. Al ser un modelo factorial, las relaciones entre factores quedan explicadas por los factores generadores de orden superior. Asimismo, se utilizó LISREL 8 para estimar ciertos índices de evaluación del ajuste. En la Tabla 11.L.8 se recogen las estimaciones paramétricas especificadas. Todas las cargas factoriales de primer nivel son estadísticamente significativas ($p < 0,01$), proporcionado validez convergente al modelo especificado. Las correlaciones múltiples al cuadrado mostraban valores adecuados. Además, todas las relaciones entre conceptos, definidas a través de las cargas factoriales de orden dos y tres, son altas y significativas. El modelo proporciona valores aceptables para los indicadores de bondad del ajuste (GFI = 0,846; AGFI = 0,824; CFI = 0,819; IFI = 0,806; AOASR = 0,048; BBNFI = 0,77). En este caso no se utiliza el índice χ^2 debido a que presenta inconvenientes de «excesiva sensibilidad» con muestras grandes, como en este caso[4]. La alternativa recomendada es el ratio χ^2/g.d.l. (Carmines y Melver, 1981) que alcanza un valor de 3,778, dentro de los límites recomendados.

7. CONCLUSIONES Y REFLEXIONES FINALES

La primera cuestión a destacar es que la confianza es un concepto complejo y multidimensional. El modelo factorial de orden tres elaborado para conceptualizar la confianza ha permitido establecer tres dimensiones básicas:

[4] La sensibilidad del índice χ^2 se debe a que se define como N-1 veces el mínimo de la función especificada para el modelo. El tamaño recomendado para evitar este problema oscila en torno a 200, considerablemente menos que la muestra utilizada (736).

TABLA 11.L.8. Estimaciones de los parámetros especificados en el modelo (estandarizadas)

Carga factorial de nivel 1		*Carga factorial de nivel 2*		*Carga factorial de nivel 3*	
Parámetro	*Valor*	*Parámetro*	*Valor*	*Parámetro*	*Valor*
V_1, F_1	0,705	F_1, F_6	0,923	F_6, F_7	0,965
V_2, F_1	0,674	F_2, F_6	0,911		
V_3, F_1	0,699	F_3, F_6	0,921		
V_4, F_1	0,761	F_4, F_7	0,503		
V_5, F_1	0,657	F_5, F_7	0,879		
V_6, F_1	0,686				
V_7, F_2	0,340				
V_8, F_2	0,392				
V_9, F_2	0,563				
V_{10}, F_2	0,556				
V_{11}, F_2	0,657				
V_{12}, F_2	0,689				
V_{13}, F_2	0,366				
V_{14}, F_3	0,523				
V_{15}, F_3	0,601				
V_{16}, F_3	0,512				
V_{17}, F_3	0,326				
V_{18}, F_4	0,830				
V_{19}, F_4	0,797				
V_{20}, F_4	0,641				
V_{21}, F_4	0,873				
V_{22}, F_4	0,661				
V_{23}, F_4	0,430				
V_{24}, F_4	0,268				
V_{25}, F_5	0,111				
V_{26}, F_5	0,365				
V_{27}, F_5	0,578				
V_{28}, F_5	0,574				
V_{29}, F_5	0,561				
V_{30}, F_5	0,531				
V_{31}, F_5	0,523				
V_{32}, F_5	0,162				
V_{33}, F_5	0,627				
V_{34}, F_6	0,299				
V_{35}, F_6	0,287				

Todas las cargas factoriales son significativas ($p < 0,01$)

1. La confianza en el vendedor, en la cual se han identificado a su vez otros tres componentes:
   1.1. Actitudes del vendedor.
   1.2. Competencia del vendedor en la función realizada.
   1.3. Intencionalidad del vendedor.
2. La relación minorista-consumidor.
3. La confianza en el establecimiento.

De nuestro análisis se extrae que la definición de confianza debe estar basada en la percepción subjetiva del consumidor final respecto a la probabilidad de ocurrencia de determinados hechos (objetivos y utilidades del consumidor final), condicionada por la confianza en el vendedor, en el establecimiento minorista y en su propia relación con el establecimiento minorista, que conducirá al mantenimiento de la relación de intercambio económico.

7.1. Implicaciones para la dirección de marketing

La contribución de nuestro trabajo al estado del conocimiento de marketing se fundamenta en tres cuestiones. En primer lugar, mientras que la mayoría de la literatura de marketing actual considera el establecimiento de relaciones como un elemento clave en la retención de consumidores, no se ha tratado con suficiente profundidad el tema de cómo dicha relación orientada a la fidelización puede ser establecida y mantenida. En éste último sentido, Barnes (1994: 563) afirma que «las relaciones de más éxito entre comprador y vendedor parecen estar caracterizadas por la *confianza mutua* entre los compradores y los vendedores, basadas en el intercambio de información, la exhibición continuada de compromiso, y un desempeño satisfactorio de los roles respectivos de las partes».

En segundo lugar, mientras que muchas organizaciones actuales están aplicando un marketing de relaciones dirigido a los consumidores finales, la literatura al respecto se ha centrado en el ámbito de las relaciones interorganizativas. Sin embargo, el establecimiento y mantenimiento de relaciones duraderas con los consumidores finales no ha sido tratado con suficiente profundidad salvo en el ámbito de los servicios. Es más, a pesar de la existencia de diversas aportaciones teóricas no existen investigaciones empíricas sobre la confianza entre minoristas y consumidores finales.

En tercer lugar, no existe un concepto de confianza universalmente aceptado, razón por la que aportamos una serie de pautas que deben configurar el concepto de confianza, derivadas del contexto de las relaciones entre minorista y consumidor final pero que pueden ser extendidas al ámbito de todas aquellas relaciones de intercambio económico donde una de las partes esté constituída por consumidores finales.

Por último, resaltar la enorme importancia de la confianza en las relaciones de intercambio debido a sus repercusiones e implicaciones tanto estratégicas como tácticas.

A nivel estratégico, sus implicaciones para la gestión empresarial son notables teniendo en cuenta la elevada intensidad de la competencia en el comercio minorista que conforma un mercado de compradores, y el enfoque relacional que el marketing está adoptando. Mientras que la perspectiva más ampliamente aceptada en la literatura actual sobre el marketing de relaciones parece ser su utilización por la parte vendedora con el resultado de crear costes a la movilidad de la parte compradora por cambio de proveedor (barreras de salida coercitivas), se puede considerar una perspectiva más amplia, más positiva y que satisfaga a ambas partes de la relación basada en la confianza (barreras de salida no coercitivas derivadas de los costes psíquicos o emocionales).

A nivel táctico, tiene implicaciones para el desarrollo de programas de fidelización de clientes y de incremento de la lealtad a la marca y/o al establecimiento.

Teniendo en cuenta los dos niveles de implicaciones citados anteriormente, la estructura multidimensional de la confianza implica que no basta con asegurar niveles aceptables de sólo algunas dimensiones, sino que todas las dimensiones analizadas y recogidas anteriormente como integrantes de la confianza deben ser satisfechas (al menos en cierto grado). Existen, por tanto, varios frentes en los que las empresas que actualmente desarrollen relaciones de intercambio con los consumidores finales deben incidir:

- Formación de los vendedores a través del proceso de selección y de la formación continuada. La actuación debe abordarse desde dos frentes: el conocimiento de las características técnicas del producto que vende y la adecuada habilidad para la comunicación interpersonal.
- Actitudes y orientación hacia el cliente de los vendedores a través de la motivación de los mismos y del establecimiento de sistemas de remuneración adecuados, así como mediante el empleo de un marketing interno.
- Creación y desarrollo de vínculos con los clientes, persiguiendo su fidelización. Se ha de fomentar una mayor frecuencia en el contacto a través de los servicios de atención al cliente y la información permanente y actualizada entre otros.
- Consideración individualizada del cliente. Esto tiene implicaciones sobre la segmentación del mercado al buscar la diferenciación para el cliente de forma que perciba una personalización en el producto-servicio
- Inversión en la mejora del capital de marca. La aceptación de sugerencias, el buen nivel de servicios, la preocupación por la calidad, la política de resolución de conflictos son factores que redundan en una mejora de la confianza. A medida que el consumidor tenga mayores niveles de confianza en el vendedor y el establecimiento, mayor será la satisfacción que tenga en el consumo de una determinada marca, incrementándose así el valor de marca.

El análisis de estos fenómenos merece una continuación en esta línea de investigación y que será objeto de estudio en trabajos posteriores.

7.2. Limitaciones

Finalmente, somos conscientes de que existen limitaciones en el estudio y análisis realizado. Por un lado, no se ha considerado el efecto que el tipo de establecimiento minorista produce sobre la percepción de la confianza por el consumidor final. Por otro lado, la confianza percibida también podría variar en función del tipo de actividad desarrollada por el establecimiento minorista. A nivel metodológico, además de los sesgos propios del proceso de selección muestral, se asumen ciertas propiedades métricas en las escalas. Asimismo, algunas de las variables consideradas como componentes de la confianza pueden ser, a su vez, consecuencias de la misma por el efecto de retroalimentación que, en la relación, produce la confianza.

BIBLIOGRAFÍA

Alba, J. y Hutchinson, W. (1987) «Dimensions of Consumer Expertise», *Journal of Consumer Research*, 13 (4), pp. 411-454.

Andaleeb, S. S. (1996) «An Experimental Investigation of Satistaction and Commitment in Marketing Channels: The Role of Trust and Dependence», *Journal of Retailing*, 72 (1), pp. 77-93.

Anderson, J. C. y Narus, J. A. (1984) «A Model of the Distributor's Perspective of Distributor-Manufacturer Working Partnerships», *Journal of Marketing*, 48 (Otoño), pp. 62-74.

Anderson, J. C. y Narus , J. A. (1990) «A Model of Distributor Firm and Manufacturer Firm Working Partnerships», *Journal of Marketing*, 54 (Enero), pp. 42-58.

Anderson, J. C. y Weitz, B. (1989) «Determinants of Continuity in Conventional Industrial Channel Dyads», *Marketing Science*, 8 (4), Otoño, pp. 310-323.

Barnes, J G. (1994) «Close to the Customer: But Is It Really a Relationship?», *Journal of Marketing Management*, 10, pp. 561-570.

Beatty, S. E., Mayer, M., Coleman, J. E., Reynolds, K. E. y Lee, J. (1996) «Customer-Sales Associate Retail Rela-

tionships», *Journal of Retailing*, 72 (3), pp. 223-247.
Bendapudi, N. y Berry, L. L. (1997) «Customers' Motivations for Mantaining Relationships With Service Providers», *Journal of Retailing*, 73 (1), pp. 15-37.
Bentler, P. M. (1995). *«EQS Structural Equation Program Manual»*. Encino: Multivariate Software, Inc.
Berry, L. L. (1995) «Relationship Marketing of Services —Growing Interest, Emerging Perspectives», *Journal of the Academy of Marketing Science*, 23 (4), pp. 236-245.
Berscheid, E. (1994) «Interpersonal Relationships», *Annual Review of Psychology*, 45, pp. 79-129.
Bitner, M. J. (1995) «Building Service Relationships: It's Al About Promises», *Journal of the Academy of Marketing Science*, 23 (4), pp. 246-251.
Blau, P. (1964) *«Exchange and Power in Social Life»*, New York, John Wiley & Sons.
Bomqvist, K. (1997) «The Many Faces of Trust», *Scandinavian Journal of Management*, 13 (3), pp. 271-286.
Bucklin, L. P. y Sengupta, S. (1993) «Organizing Successful Co-Marketing Alliances», *Journal of Marketing*, 57 (2), pp. 32-46.
Carmines, E. G. y McIver, J. P. (1981). «Analyzing Models with Unobserved Variables: Analysis of Covariances Structures», en Bohrnstedt, G. W. y Borgatta, E. F. (eds.) *Social Measurement: Current Issues*. Beverly Hills: Sage, pp. 65-115.
Cespedes, F. V. y Smith, H. J. (1993) «Database Marketing: New Rules for Policy and Practice», *Sloan Management Review*, 35 (Verano), pp. 7-22.
Clopton, S. W. (1984) «Seller and Buying Firm Factors Affecting Industrial Buyer's Negociation Behavior Outcomes», *Journal of Marketing Research*, 21 (Febrero), pp. 39-53.
Crosby, L. A., Evans, K. R. y Cowles, D. (1990) «Relationship Quality in Services Selling: An Interpersonal Influence Perspective», *Journal of Marketing*, 54 (Julio), pp. 68-81.
Childers, T. L. y Rao, A. R. (1992) «Influence of Familial and Peer-Based Reference Groups on Consumer Decisions», *Journal of Consumer Research*, 19, septiembre, pp. 198-211.
Churchill, G. A. (1979). «A Paradigm for Developing Better Measures of Marketing Constructs», *Journal of Marketing Research*, Vol. 16 (Febrero), pp. 68-73.
Dant, R. P. y Schul, P. L. (1992) «Conflict Resolution Processes in Contractual Channels of Distribution», *Journal of Marketing*, 56 (Enero), pp. 38-54.
Deutsch, M. (1960) «The Effect of Motivational Orientation on Trust and Suspicion», *Human Relations*, 13, pp. 123-139.
Doney, P. M. y Cannon, J. P. (1997) «An Examination of the Nature of Trust in Buyer-Seller Relationships», *Journal of Marketing*, 61 (Abril), pp. 35-51.
Dwyer, F. R., Schurr, P. H. y Oh, S. (1987) «Developing Buyer-Seller Relationships», *Journal of Marketing*, 51 (Abril), pp. 11-27.
Fornell, C. y Larcker, D. F. (1981). «Evaluating Structural Equations Models with Unobservable Variables and Measurement Error», *Journal of Marketing Research*, Vol. 18 (Febrero), pp. 39-50.
Frazier, G. L. (1983) «On the Measurement of Interfirm Power in Channels of Distribution», *Journal of Marketing Research*, 20 (Mayo), pp. 158-166.
Gambetta, D. (1988) «Can We Trust Trust?» en D. Gambetta (Ed.) *Trust: Making and Breaking Co-operative Relations*, Oxford: Basil Blackwell.
Ganesan, S. (1994) «Determinants of Long-Term Orientation in Buyer-Seller Relationships», *Journal of Marketing*, 58 (Abril), pp. 1-19.
Grönroos, C. (1989) «Defining Marketing: A Market-Oriented Approach», *European Journal of Marketing*, 23 (1), pp. 52-60.
Grönroos, C. (1990) «Relationship Approach to Marketing in Service Contexts: The Marketing and Organisational Behaviour Interface», *Journal of Business Research*, 20 (Enero), pp. 3-11.
Grönroos, C. (1994) «From Marketing-Mix to Relationship Marketing: Towards a Paradigm Shift in Marketing», *Management Decision*, 32 (2), pp. 4-20.
Grönroos, C. (1996) «Relationship Marketing: Strategic and Tactical Implications», *Management Decision*, 34 (3), pp. 5-14.
Gundlach, G. T. y Murphy, P. E. (1993) «Ethical and Legal Foundations of Relational Marketing Exchanges», *Journal of Marketing*, 57 (Abril), pp. 35-46.
Gurviez, P. (1997) «Trust: A New Approach to Understanding the Brand-Consumer Relationship», *Proceedings of the American Marketing Association (AMA)*, Dublin, Junio, pp. 504-518.
Håkansson, K. (1982) «An Interaction Approach», H. Håkansson (Ed.) *International Marketing and Purchasing of Industrial Goods: An Interaction Approach*, Chichester: John Wiley & Sons, pp. 10-27.
Hallén, L, Johanson, J. y Seyed-Mohamed, N. (1991) «Interfirm Adaptation in Business Relationships», *Journal of Ma rketing*, 55 (Abril), pp. 29-37.
John, G. y Reve, T. (1982). «The Reliability and Vadility of Key Informant Data from Dyadic Relationships in Marketing Channels», *Journal of Marketing Research*, Vol. 19 (Noviembre), pp. 517-524.
John, G. (1984) «An Empirical Investigation of Some Antecedents of Opportunism in a Marketing Channel», *Journal of Marketing Research*, 21 (Agosto), pp. 279-289.
Jöreskog, K. G. y Sörbom, D. (1993). *«LISREL 8 User's Reference Guide»*. Chicago: Scientific Software International.
Kohli, A. K. y Jaworski, B. J. (1990) «Market Orientation: The Construct, Research Propositions and Management Implications», *Journal of Marketing*, 54 (Abril), pp. 1-18.
Kotler, P., Cámara, D. y Grande, I. (1995) *«Dirección de Marketing»*, 8ª ed., Madrid: Prentice Hall.
Macaulay, S. (1963) «Non-contractual Relations in Business: A Preliminary Study», *American Sociological Review*, 28 (Febrero), pp.55-69.
Macneil, I. R. (1978) *«The New Social Contract. An Inquiry into Modern Contractual Relations»*, New Haven: Yale University Press.

Madhok, A. (1995) «Opportunism and Trust in Joint Venture Relationships: An Exploratory Study and a Model», *Scandinavian Journal of Management*, 11 (1), pp. 57-74.

Moorman, C., Deshpandé, R. y Zaltman, G. (1993) «Factors Affecting Trust in Market Research Relationships», *Journal of Marketing*, 57 (Enero), pp. 81-101.

Moorman, C., Zaltman, G. y Deshpandé, R. (1992) «Relationships Between Providers and Users of Market Research: The Dynamics of Trust Within and Between Organizations», *Journal of Marketing Research*, 29 (Agosto), pp. 314-329.

Morgan, R. M. y Hunt, S. D. (1994) «The Commitment-Trust Theory of Relationship Marketing», *Journal of Marketing*, 58 (Julio), pp. 20-38.

Murphy, P. E. y Gundlach, G. T. (1997) «A Typology of Trust in Business», *Proceedings of the American Marketing Association Special Conferences*, Dublín, pp. 596-598.

Nunnally, J. C. (1987). *«Teoría Psicométrica»*. México: Trillas.

Parasuraman, A., Berry, L. L. y Zeithaml, V. A. (1991) «Understanding Customer Expectations of Service», *Sloan Management Review*, Primavera, pp. 39-48.

Parasuraman, A., Zeithaml, V. A. y Berry, L.L. (1985) «A Conceptual Model of Service Quality and Its Implications for Future Research», *Journal of Marketing*, 49 (Otoño), pp. 41-50.

Peterson, R. A. (1994). «A Meta-analysis of Cronbach's Coefficient Alpha», *Journal of Consumer Research*, Vol. 21 (Septiembre), pp. 381-391.

Pfeffer, J. y Salancick, G. (1978) *«The External Control of Organizations: A Resource Dependence Perspective»*, Nueva York: Harper & Row.

Ravald, A. y Grönroos, C. (1996) «The Value Concept and Relationship Marketing», *European Journal of Marketing*, 30 (1), pp. 19-30.

Rich, G. A. (1997) «The Sales Manager as a Role Model: Effect of Trust, Job Satisfaction and Performance of Salespeople», *Journal of the Academy of Marketing Science*, 25 (4), pp. 319-328.

Rotter, J. B. (1967) «A New Scale for the Measurement of Interpersonal Trust», *Journal of Personality*, 35 (4), pp. 651-665.

Schurr, P. H. y Ozanne, J. L. (1985) «Influences on Exchange Processes: Buyer's Preconceptions of a Seller's Trustworthiness and Bargaining Toughness», *Journal of Consumer Research*, 11 (Marzo) pp. 939-953.

Shapiro. S. P. (1987) «The Social Control of Impersonal Trust», *American Journal of Sociology*, 93 (Noviembre), pp. 623-658.

Smith, J. B. y Barclay, D. W. (1997) «The Effects of Organizational Differences and Trust on the Effectiveness of Selling Partner Relationships», *Journal of Marketing*, 61 (Enero), pp. 3-21.

Stern, L. W. y El-Ansary, A. I. (1992) *Marketing Channels*, 4ª ed., Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Webster, F.E. (1992): «The Changing Role of Marketing in the Corporation», *Journal of Marketing*, 56 (Octubre), pp. 1-17.

Werts, C. E.; Rock, D. R.; Linn, R. L. y Jöreskog, K. G. (1978). «A General Method of Estimating the Reliability of a Composite», *Educational and Psychological Measurement*, Vol. 38, pp. 933-938.

Wilson, D. T. (1995) «An Integrated Model of Buyer-Seller Relationships», *Journal of the Academy of Marketing Science*, 23 (4), pp. 335-345.

Williamson, O. E. (1975) *«Markets and Hierarchies: Analysis and Antitrust Implictions»*, Nueva York: The Free Press.

Williamson, O. E. (1993) «Calculativeness, Trust, and Economic Organization», *Journal of Law & Economics*, 36, abril, pp. 453-486.

CAPÍTULO 12

Nuevas técnicas de análisis multivariante

OBJETIVOS DE APRENDIZAJE

- Entender los principios subyacentes del almacenamiento de datos.
- Identificar y describir las técnicas asociadas de la extracción de datos.
- Describir los principios específicos de las redes neuronales.
- Explicar las operaciones básicas de las redes neuronales, especialmente los procesos de aprendizaje.
- Distinguir entre los métodos paramétricos y de reespecificación de estimación de parámetros.
- Ennumerar las ventajas de los métodos de bootstrap y jackknife.

La transición de una economía basada en la producción a una basada en la información ha tenido un profundo efecto no sólo en las operaciones de las organizaciones, sino también en el entorno del análisis de los negocios. Los investigadores se enfrentan a nuevos retos que requieren el uso de las técnicas tradicionales multivariantes de diferentes modos y que combinan estas técnicas tradicionales con nuevas técnicas analíticas. La tremenda cantidad de información generada actualmente por las organizaciones sólo puede aumentar. A medida que el comercio y las comunicaciones electrónicas se han hecho más comunes en Internet, las empresas están desarrollando un vínculo directo con los clientes, otra fuente de información.

El investigador de hoy día debe confrontar las tareas de la gestión y el análisis de los datos. La gestión de los datos está beneficiándose del concepto de almacenamiento de datos, que integra datos en información apropiada para las aplicaciones analíticas. El análisis de datos se ha visto beneficiado por la existencia de dos nuevos procedimientos que complementan a las técnicas

tradicionales del análisis multivariante. El primero es el modelo de aprendizaje, tipificado por redes neuronales, que se ajusta particularmente para el análisis de pautas complejas sin una especificación previa del modelo. El segundo es la reespecificación, que reemplaza las bases estadísticas de la inferencia con una estimación estrictamente empírica basada es muestreos repetidos de la muestra original. El poder de cálculo disponible hoy en día permite al investigador estimar empíricamente el intervalo de confianza de un parámetro en lugar de basarse en supuestos estadísticos. Este capítulo contiene una breve introducción a cada uno de estos temas, así como ejemplos empíricos donde proceda.

Introducción

Hasta hace poco, el carácter fundamental de los análisis de datos ha permanecido básicamente inmóvil desde el último gran desplazamiento de mediados de los 70, cuando la introducción de los computadores y su evolución a computadores personales provocó un cambio revolucionario en la práctica del análisis de datos. Desde entonces, los investigadores han dispuesto de un poder de cálculo suficiente como para eliminar prácticamente todas las restricciones de los diferentes tipos de técnicas estadísticas existentes, con lo que los métodos estadísticos multivariantes que hemos discutido en los capítulos anteriores han ganado amplia aceptación y uso. El peso de la educación del investigador se ponía en los métodos estadísticos, con énfasis en los fundamentos teóricos y aplicaciones de las técnicas multivariantes. Se consideraba poco la naturaleza o el carácter de los datos, salvo la evaluación de las calidades estadísticas en términos de cumplimiento de los supuestos de las técnicas estadísticas.

Pero la «era de la información» ha traído una segunda revolución, particularmente fuera del ámbito académico. De nuevo, los investigadores se ven desafiados, esta vez reorientados a la aplicación de sus técnicas para un nuevo entorno de investigación. Las bases de datos a gran escala con cientos de miles o incluso millones de observaciones ofrecen unos perfiles de atributos detallados durante lapsos de tiempo que pueden durar años. Esta explosión de datos contrasta con el pasado, donde el investigador a menudo analizaba datos recogidos específicamente para la investigación en cuestión. Ahora los datos abundan, y las organizaciones buscan nuevas formas de extraer información. Esta segunda revolución en el análisis de datos se basa fundamentalmente en dos tendencias: una avalancha de información y cuestionar la inferencia estadística.

La avalancha de información

Tanto en las comunidades comerciales como en las académicas, los días de muestras limitadas o tamaños muestrales reducidos son historia. El mundo académico tiene acceso a bases de datos publicadas con temas que van desde el control de datos de las compras de los consumidores a los registros históricos de información financiera completa de todas las empresas registradas. Junto con los tipos de información emergentes disponibles en fuentes gubernamentales, los académicos tienen en muchos casos datos más que suficientes. Pero en el sector comercial, el escenario es mucho más complicado. Organizaciones de todo tipo se han ido informatizando cada vez más en todas sus áreas funcionales, facilitando no sólo sus operaciones, sino también su habilidad para recoger datos. Los avances tecnológicos ofrecen la capacidad de almacenar efectivamente toda la información en un formato común, y el acceso a los datos se encuentra disponible para todos los miembros de la organización, no sólo para unos pocos elegidos. Estas tendencias se han combinado para crear algo parecido a una «avalancha de la información», mientras que el investigador puede sentirse fácilmente abrumado por el flujo continuo de cantidades masivas de datos. Bases de datos de cien *gigabytes* son habituales, y ya se empieza a pensar en bases de datos de un *terabyte* (1.000 *gigabytes*). ¿Cómo puede procesarse esta información y analizarse desde una perspectiva temporal? ¿Qué técnicas son las más apropiadas para este nuevo desafío? Éstos son algunos de los temas a que se enfrentan los investigadores en este nuevo entorno de investigación.

Análisis sin inferencia estadística

La avalancha de información también ha prestado atención a las alternativas a la inferencia estadística para evaluar la «significación» de las estimaciones de los parámetros. Pero, ¿por qué querríamos abandonar este principio –la base de la mayoría de las técnicas multivariantes discutidas en este texto? En primer lugar, hay un movimiento a favor de «volver a los datos», en el que el investigador aplica los menores supuestos posibles para el análisis y «deja hablar a los datos». El desarrollo de los programas de inteligencia artificial y su progenie (es decir, redes neuronales y algoritmos genéticos) ha llevado a los investigadores a explorar un amplio rango de modelos analíticos que no están fundamentados en inferencias estadísticas. Estas técnicas se basan en la aplicación de reglas simples de aprendizaje a un conjunto de datos de la forma menos restringida posible. Muchos investigadores vieron inicialmente estos modelos como una «caja negra», en las cuales se produce algún proceso misterioso que da lugar a unos resultados. Esto está lejos de ser verdad porque estos métodos están muy estructurados en sus procedimientos —sólo que reaccionan a los datos en lugar de que los datos se perfilen para que se ajusten a los supuestos estadísticos.

En segundo lugar, la comunidad académica ha hecho un llamamiento para enseñar a los estudiantes los fundamentos básicos de la estadística sin la memorización de fórmulas y teoría estadística. Los estudiantes podrán asimilar fácilmente los conceptos de la inferencia estadística cuando puedan visualizar realmente el proceso, como demostrar el concepto de una distribución de probabilidad mediante lanzamiento o combinaciones de monedas. Así que, ¿por qué descartar esta sencilla perspectiva cuando vamos hacia técnicas más sofisticadas? Combinado con el poder de cálculo que existe hoy en día, estos métodos permiten una estimación empírica de la distribución en lugar de un supuesto estadístico. Esto es parecido a lanzar la moneda al aire muchas veces, pero ahora con cuestiones y técnicas analíticas decididamente más complejas. Todo lo anterior ha llevado a la incorporación de estas alternativas, tales como redes neuronales y técnicas de reespecificación a la caja de herramientas del investigador.

Temas abarcados en este capítulo

Los cambios y fuerzas en nuestro entorno de investigación se han evidenciado de muchas maneras en los últimos años, emergiendo tres conceptos específicos, como representantes de nuestro claro abandono de nuestros antiguos modos de pensar sobre el análisis de datos: el almacenamiento de datos y su extracción, las redes neuronales y la reespecificación. El propósito de este capítulo no es ofrecer una discusión en profundidad de cada tema, sino discutir los conceptos básicos, la terminología y los supuestos, utilizando como punto de referencia la exposición a que ha estado sometido el lector en este texto respecto a las técnicas estadísticas multivariantes. La siguiente sección contiene unas breves discusiones de cada tópico, que serán seguidas de una cobertura más detallada con ejemplos empíricos donde se puedan aplicar. Se anima al lector a revisar en primer lugar estas descripciones, para obtener un conocimiento básico de los conceptos, y a continuación revisar su provecho en situaciones específicas.

*Almacenamiento y extracción de datos**

Ningún concepto ha generado más atención en la comunidad de la tecnología de la información (IT) como el almacenamiento de datos, y después la extracción de datos [12, 14]. El **almacenamiento de datos** es el intento de combinar todas las fuentes de información y datos relevantes para una organización en una única base de datos con una estructura susceptible de apoyar el proceso

* Hemos decidido traducir «Data Mining» como extracción de datos y así denominaremos a este proceso durante el resto del capítulo, aunque en algunas ocasiones la literatura académica o los investigadores de mercado no lo hayan traducido y utilicen la expresión anglosajona.

analítico de toma de decisiones a todos los niveles de la organización. Existen muchos supuestos de una naturaleza muy técnica que no se discutirán aquí, pero es esencial que el investigador utilice el almacenamiento de datos para entender su estructura básica y operación en la medida en que manifiesta su utilidad para varios tipos de análisis. La **extracción de datos** es en cierta forma una nueva perspectiva del análisis de datos con más orientación exploratoria que confirmatoria. Dada la enorme cantidad de datos disponibles, la perspectiva del investigador se dirige más hacia la profundización en las características de los datos y no centrarse en la generalización a otras situaciones. Por tanto, se emplean otras técnicas además de los métodos multivariantes tradicionales, muchas veces de una naturaleza simple y cualitativa. El investigador, dada su profundidad de la información, sigue una estrategia de descubrimiento a través del examen de los datos para todos los tipos de relaciones.

Redes neuronales

Aunque inicialmente se considerasen como un método utilizado por los científicos de la computación en sus esfuerzos en la **inteligencia artificial**, las redes neuronales se han convertido en una herramienta aceptada por los investigadores de a pie, particularmente por aquellos ajenos al mundo académico. Siguiendo el ímpetu del almacenamiento de datos y extracción de datos, las redes neuronales ofrecen una aproximación relativamente directa de exploración y descubrimiento [4, 22, 33]. También, dada la naturaleza más pragmática de esta exploración, las redes neuronales ofrecen una poderosa herramienta de predicción basada en la cuantificación y replicación de pautas complejas en los datos. Se hace menos énfasis en lo que se está haciendo y más en cómo se está haciendo. Aunque las redes neuronales pueden replicar y muchas veces sobrepasar a muchas técnicas estadísticas, como la regresión múltiple, análisis discriminante, regresión logística y análisis cluster, no se interpreta fácilmente debido a las complejas relaciones que se manejan de forma «invisible» por la metología. Por tanto, se anima al investigador a emplear redes neuronales cuando la exploración y la predicción, pero no la explicación, sean los puntos principales de la investigación.

Remuestreo

El punto final —remuestreo— podría sorprender a muchos por su inclusión en un texto de análisis multivariante. Pero el remuestreo, la aproximación analítica que sustituye la estimación empírica de distribuciones muestrales por las propiedades supuestas a distribuciones teóricas, como las distribuciones normales o *t*, difiere sólo en sus supuestos subyacentes sobre los datos y sus características. Todas las técnicas multivariantes caracterizadas por la inferencia estadística clásica, como la regresión múltiple, el análisis discriminante o MANOVA, pueden desarrollarse fácilmente con el remuestreo. La capacidad de cálculo disponible hoy en día permite al usuario evitar los supuestos de normalidad mediante la extracción repetida de muestras y definir empíricamente la distribución muestral efectiva. De esta forma, el investigador está de nuevo «volviendo a los datos» y representando las características efectivas de los datos en lugar de asumir ciertas características a efectos de la estimación. Como con los temas ya discutidos, muchos de los principios y técnicas que hemos aprendido siguen siendo operativos; sólo estamos utilizando procesos algo distintos para la estimación, inferencia e interpretación del modelo.

Almacenamiento y extracción de datos

La emergencia del almacenamiento y extracción de datos puede describirse mejor por el dicho «la necesidad es la madre de la invención». Las organizaciones han apoyado de forma entusiasta los aumentos en la automatización e informatización a lo largo de los años y valorado el papel estra-

tégico de la información. Pero se han combinado tres fuerzas para forzar la comprensión de que los nuevos enfoques eran necesarios para la asimilación y uso efectivo de la información disponible en las actuales economías intensivas en información [2, 28].

El primer factor es la expansión del conjunto de información que casi todas las organizaciones tienen a su disposición. Los sistemas de contabilidad informatizados hacen que la información efectiva se encuentre rápidamente disponible en todos los niveles de la organización. La automatización de los procesos de fabricación, particularmente la obtención y aumento de los sistemas de control de inventarios, también ofrece una continua fuente de información. La participación en una economía global y la necesidad de operaciones internacionales exige incluso más cooperación y control, una nueva razón para aumentar la informatización. La reciente aparición de Internet sólo promete un aumento de la oferta y demanda de información, particularmente a medida que el comercio electrónico ofrece un vínculo directo a los consumidores e información que hasta ahora sólo estaba disponible para los detallistas. La automatización de la fuerza de ventas, fundamentalmente en la informatización y una «oficina virtual», ofrece más de una fuente de información que hasta hace poco se encontraba en un formato que no beneficiaba a nadie salvo al vendedor. Las organizaciones han evolucionado desde basarse estrictamente en datos operativos a la integración de información externa dirigida hacia una función específica, sea la interacción de la satisfacción del cliente para el departamento de marketing o los informes financieros para evaluar la evolución del mercado y su rendimiento.

Una segunda fuerza relacionada es que muchas organizaciones, como consecuencia del proceso descrito previamente, están produciendo información tan rápido como productos o servicios. Las organizaciones, al capturar los datos del resto de los procesos, están creando un papel paralelo: evaluación constante e inmediata de sus operaciones. Esta fuerza tropezó con la tercera —técnicas que pueden ajustarse a la producción de información con la producción de conocimiento. Lo que se necesitaban eran análisis que se modificaban sin intervención humana, liberando al investigador de centrarse en los resultados y el rendimiento del modelo. A medida que estas técnicas se hicieron asequibles y aceptadas, las tres fuerzas se alinearon para promover la acción. Hemos asistido últimamente a una explosión en el uso y aplicación del almacenamiento y extracción de datos. Las siguientes secciones detallan las características principales tanto del almacenamiento como de la extracción de datos, finalizando con una discusión de las implicaciones que tiene para las técnicas multivariantes.

¿En qué consisten la extracción y el almacenamiento de datos?

La extracción y almacenamiento de datos son elementos complementarios en la mejora del acceso a los datos para la toma de decisiones. El almacenamiento de datos es el mecanismo que facilita el **sistema de apoyo a la decisión (DSS)**, almacenando los datos de la organización en una única base de datos integrada y ofrece una perspectiva histórica. Los dos conceptos claves que subyacen en el almacenamiento de datos son la integración y la invarianza en el tiempo [18, 20]. La integración se refiere al diseño unificado de las bases de datos que combinan todas las fuentes de la organización en un único punto de acceso. La invarianza en el tiempo significa que preserva una perspectiva histórica, de tal forma que tales «trozos de realidad» se encuentren disponibles para cualquier análisis retrospectivo. Un almacenamiento de datos, sin embargo, no es una aplicación, sino que facilita las aplicaciones ofreciendo datos desde una perspectiva temporal en el formato adecuado. Su papel como inversión estratégica es cada vez más aceptado [6, 28].

La extracción de datos, también conocida como **descubrimiento de conocimiento en las bases de datos (KDD)**, es la búsqueda de relaciones y pautas entre los datos en grandes bases de datos [3, 12, 13, 26]. Como sugiere el término, la extracción de datos tiene una orientación exploratoria de búsqueda de conocimiento oscurecido por las pautas de asociación complejas y la gran cantidad de datos. Sólo se pueden encontrar unas pocas «pepitas de oro» de conocimiento después de procesar enormes cantidades de información. Un supuesto implícito es que mediante la revelación de estas relaciones en las bases de datos, los beneficios aumentarán en la medida en que la base

de datos refleje perfectamente el entorno de decisión de la organización. Se pueden emplear un gran número de técnicas analíticas en la extracción de datos, que van desde los enfoques gráficos y descriptivos más básicos a las más sofisticadas técnicas multivariantes (es decir, análisis cluster o regresión múltiple o logística) y los novísimos modelos de aprendizaje (es decir, redes neuronales y algoritmos genéticos) [3, 33].

Conceptos fundamentales del almacenamiento de datos

El principal objetivo del almacenamiento de datos es la integración de datos de toda la empresa en un formato asequible al análisis. Por tanto, el almacenamiento de datos consigue al menos tres objetivos: (1) proporciona apoyo para el DSS en términos de acceso y organización de los datos; (2) segrega el acceso al almacén de datos, reduciendo por tanto la degradación de la evolución de los sistemas operacionales debido a las peticiones repetidas; y (3) fuerza a un reconocimiento de las diferentes estructuras de datos necesarias a efectos analíticos frente a los operativos.

Operando en un almacén de datos

La operación de un almacén de datos tiene muchos aspectos técnicos sofisticados concernientes a la organización de los datos, estructuras de datos y diseño de las bases de datos que están fuera del alcance de este texto. Los lectores interesados pueden desear revisar un texto introductorio de diseño y arquitectura de un almacén de datos [8, 17, 18, 21, 31, 34]. Desde una perspectiva operativa, el almacenamiento de datos implica siete fases básicas:

1. *Adquisición de datos*. La primera tarea es adquirir datos de todas las fuentes relevantes, tanto externas como internas a la organización [2]. Las fuentes de datos internas son el principal legado de los sistemas (existentes) que tratan con los procesos operativos, tales como los procesos de transacción y gestión de inventarios. Las fuentes de información externa pueden consistir en información del cliente (es decir, datos demográficos), estudios de investigación (es decir, estudios de satisfacción u otros resultados de la investigación del marketing), o incluso fuentes comerciales, tales como información de censos o bases de datos comerciales (es decir, perfiles Dunn y Bradstreet). Ninguno de estos datos se encuentra en diferentes niveles de análisis o agregación y deben integrarse con otra información al mismo nivel de agregación.
2. *Integración de datos*. Esta fase integra todas las fuentes de datos, manteniendo la consistencia mediante la combinación de características, atributos y nivel de agregación. Como discutiremos en una segunda sección, los datos adquieren su calidad multidimensional durante esta fase.
3. **Limpieza de datos**. La integración de muchas fuentes de datos necesita un escrutinio de los datos para eliminar errores y la baja calidad de los datos, mientras a la vez se llevan a cabo controles de consistencia [27]. La calidad de todos estos análisis depende de la calidad subyacente de los datos, convirtiéndose en uno de los mayores obstáculos en la puesta en práctica con éxito del almacenamiento de los datos.
4. *Creación de metadatos*. Los metadatos comprenden una descripción completa de un **elemento de datos**, que incluye no sólo sus atributos, sino también su fuente original así como cualquier resumen o transformación. La descripción de metadatos proporciona un perfil de los elementos de datos que pueden utilizarse en un análisis, transformación o resumen ulterior.
5. *Importación de datos*. A intervalos periódicos, los datos se importan a un almacén de datos. Dado que éstos son en realidad un archivo histórico, estos datos se introducen en la base de datos existente y ofrecen una perspectiva temporal.
6. *Almacenamiento de datos*. Esta fase incluye la organización y procesamiento de la base de datos, así como el resumen de los datos a varios niveles para anticiparse a las peticiones del usuario.

7. *Apoyo a la decisión.* El apoyo a la decisión se refiere a las solicitudes dirigidas por los usuarios desde **OLAP (procesamiento de datos en línea)** o procedimientos de extracción de datos que acceden a una base de datos desde una petición especificada. OLAP comprende aquellas peticiones en «tiempo real» de la base de datos que acceden a resúmenes de datos multidimensionales preespecificados, para obtener una respuesta inmediata cuando sea posible[31]. Las técnicas de extracción de datos se especializan en procedimientos analíticos, que serán discutidos con más detalle en la sección siguiente.

Al describir las operaciones del almacenamiento de datos, es importante distinguir entre dos sistemas: **sistemas operacionales** y sistemas de toma de decisiones (DSS). Un sistema operacional controla las funciones básicas **(transacciones)** de los negocios, tales como contabilidad, existencias y gestión de pedidos. Sin estos sistemas, la organización no podría funcionar. Los sistemas operacionales exigen datos que reflejen el status actual de la organización; los datos históricos de escaso interés. Como contraste, los sistemas de apoyo a la decisión implican solicitudes que tratan con la planificación de la estrategia de la organización [30, 33]. A corto plazo, una organización podría existir sin el DSS. Pero dada la naturaleza crítica de la información, que se constituye como una capacidad de la organización para competir y reaccionar al entorno actual de los negocios, una organización sin el DSS tendrá vida corta. Los sistemas operacionales funcionan con sus propias bases de datos sobre la base de la eficacia del procesamiento de múltiples peticiones individuales. Un sistema de apoyo a la decisión, sin embargo, exige la existencia de un almacén de datos para recoger y convertir los datos a un formato adecuado para las solicitudes DSS. Sin un almacén de datos, se gasta una enorme cantidad de tiempo y esfuerzo en la preparación de los datos para cada petición de información al DSS.

Definiciones de los datos

Los beneficios de un almacén de datos no sólo provienen de sus operaciones día a día y del almacenamiento de datos, sino también de la definición de los datos. Tanto si se definen sus características básicas o nivel de agregación, como si se forma un perfil completo, estas definiciones de datos son instrumentales en la eficacia y efectividad del almacenamiento de los datos. Las tres definiciones más importantes son: datos operacionales frente a analíticos, datos originarios frente a agregados y metadatos.

Datos operacionales frente a datos analíticos Los datos operacionales son el fundamento de mantenimiento automatizado de las operaciones día a día de la organización, pero no se ajustan bien a un análisis más profundo. Un interés central del almacenamiento de los datos es la creación de los datos analíticos mediante el procesamiento e integración de datos operativos en la base de datos en un formato adecuado a las solicitudes de extracción de datos [18, 19]. Los datos analíticos reflejan una estructura interna que combina eventos separados alrededor de un objeto común, tanto si se trata de la historia de las compras de un cliente como las ventas de un producto u otra unidad del análisis. Estos procesos de transformación ponen de manifiesto que el almacenamiento de los datos no sólo consiste en el depósito de datos, sino también en cómo se almacenan.

Datos originales frente a datos agregados Una segunda característica de un almacenamiento de datos es su capacidad para resumir **datos originales** y almacenarlos como **datos agregados** a un nivel superior de análisis. Los sistemas operacionales descansan en los datos originales dado que se centran sólo en eventos únicos, no en compuestos de eventos. Las aplicaciones de DSS, sin embargo, pueden beneficiarse del aumento de la velocidad de acceso en la oferta de datos agregados. Este asunto es crucial a medida que la base de datos aumenta de tamaño, haciendo que las peticiones que requieren resúmenes de datos sean menos eficientes. Además, la rápida respuesta de OLAP se desprende de su acceso a los datos agregados, evitando los retardos computacionales asociados con el proceso de resumirlos. La clave está en agregar sólo

la información que se necesita, dado que los datos agregados pueden incorporarse en gran medida a los requisitos de almacenamiento. A medida que la dimensión de la base de datos y la complejidad aumentan, el número de posibles agregaciones excede rápidamente cualquier límite factible, de tal forma que el personal de almacenamiento de datos debe seleccionar las agregaciones que ellos creen satisfacen mejor a sus usuarios.

Esto no significa, sin embargo, que los datos primitivos sean útiles sólo de forma agregada. La retención de los datos originales permite el proceso de **búsqueda concreta**, ya que el investigador puede querer examinar los datos que subyacen en los datos agregados seleccionados. Suponga que después de revisar los totales de ventas mensuales, el investigador quiere ver las ventas totales por punto de venta y por semana y a continuación se centra en las ventas de una tienda particular por línea de producto. La retención de los datos originales permite la agregación a cualquier nivel y sobre cualquier dimensión. Esto facilita el proceso de descubrimiento mediante la exploración de la manera más flexible posible.

Metadatos La definición final de datos que caracteriza el almacenamiento de los datos es la de **metadatos**. Literalmente «datos sobre los datos», los metadatos ofrecen un perfil completo de los elementos de datos, incluyendo su fuente, transformaciones, cualquier resumen, una lista completa de dimensiones, plazo temporal y cualquier otra información pertinente [2, 18, 19, 20]. Los metadatos también permiten un sistema de clasificación estándar entre los elementos de los datos. Por ejemplo, el término ventas tiene muchos significados en una organización. Para el marketing, una transacción con clientes. Para los gestores de las existencias, es un producto que se tiene que reemplazar. Para el departamento de contabilidad, ingresos menos devengos y descuentos. Cada una de estas cosas son ventas dentro de un área funcional, pero difieren ligeramente en su marco temporal como cuando se producen y las acciones que exigen. Los metadatos permiten una clasificación uniforme que especifica el carácter distintivo de cada elemento de los datos en el almacén de datos.

Resumen

El almacén de datos, en concepto, es una idea que ha rondado las cabezas de los investigadores que han manejado datos dentro de las organizaciones. Cualquier persona con experiencia en sistemas de bases de datos de organizaciones le dirá que rara vez los datos están estructurados en un formato que ofrezca el tipo de información necesaria para los sistemas de apoyo a la decisión. El almacenamiento de datos surge como respuesta a las enormes cantidades de datos que se acumulan en organizaciones y la necesidad de un proceso sistemático de traslación de datos operacionales a datos analíticos. Su continuo desarrollo fomenta no sólo el uso de aplicaciones DSS, sino también el reconocimiento del valor estratégico implicado en la información de la base de datos de la organización.

Supuestos fundamentales en la extracción de datos

Una vez que la base de datos está en su sitio, el investigador se encuentra preparado para el siguiente paso: la extracción de datos. Como ya se ha descrito, la extracción de datos es el proceso de extracción de información de grandes bases de datos. Dada su naturaleza analítica, discutiremos primero algunas diferencias y parecidos generales con el análisis multivariante tratadas en capítulos anteriores. Una segunda discusión se centra en la orientación predominante de la extracción de datos —exploración de datos— y cómo ésta cambia el enfoque del investigador respecto del análisis de los datos. A continuación, la discusión se centra en los tipos básicos de técnicas analíticas utilizadas en la extracción de datos que van desde las formas más simples (peticiones) a las más complejas (redes neuronales y algoritmos genéticos). Concluimos examinando algunas recomendaciones específicas del investigador al hacer la transición de la orientación basada en la estadística a esta orientación diferente hacia el análisis de los datos.

Diferencias y similitudes en la extracción de datos

El investigador notará que la extracción de datos no es una técnica analítica sino una aproximación al análisis de los datos. Como tal, se aplica mucho de lo que hemos tratado en capítulos precedentes. Pero la extracción de datos también tiene ciertas características únicas, especialmente en su aproximación al proceso analítico en general.

Algunas diferencias Hasta este momento, hemos tratado con problemas de investigación bien definidos para los que se ha desarrollado un plan de investigación detallado, incluyendo la selección de la técnica multivariante apropiada. Muy a menudo se utiliza un modo confirmatorio, en el que los modelos propuestos se evalúan con su explicación de la situación. La extracción de datos, sin embargo, presenta una situación algo diferente por varias razones.

En primer lugar, muchas veces el investigador está operando en un modo exploratorio estructurado alrededor de una cuestión de investigación muy general (es decir, ¿qué afecta a la retención del cliente?). También, aunque la extracción de datos acompaña a un amplio rango de herramientas analíticas, tenemos menos guía en la selección de una técnica de la que hemos tenido en nuestras anteriores discusiones, en las que la escala de medida y el número de variables para los valores teóricos dependientes/independientes determinan el método que se va a emplear. Dado que la exploración es el principal punto focal de la extracción de datos, muchas de las técnicas, necesariamente, están muy generalizadas y pueden emplearse para cualquier número de funciones diferentes y para todos los tipos de variables. Por tanto, el investigador selecciona el método analítico de su aproximación al problema, así como su sofisticación analítica o los requisitos explícitos para la entrada de datos.

Algunas similitudes Muchos de los temas a los que nos hemos enfrentado en los capítulos anteriores, surgen cuando discutimos la extracción de datos. Incluso aunque difieren en forma muchas veces, la extracción de datos se basa en análisis multivariantes, incluso con los métodos más simples diseñados para utilizar un número amplio de variables. La naturaleza del concepto de almacenamiento de los datos implica una verdadera perspectiva multivariante. Quizá más que en ningún otro momento, la teoría puede jugar un papel importante en el éxito del esfuerzo de extracción de los datos. Las grandes bases de datos utilizadas para la extracción de datos no sólo contienen muchas observaciones, sino que a menudo también incluyen una cantidad tremenda de información sobre cada encuestado. Sin embargo, sin guía, el proceso exploratorio puede abrumar al analista, mientras que el confirmatorio incluso no es posible. Por tanto, el investigador nunca debe «dejar pensar a la máquina» en términos de formulación de diseño de investigación. Algunos métodos, como las redes neuronales u otros métodos de aprendizaje, pueden requerir una escasa especificación de las relaciones efectivas, pero sufrirán tanto como cualquier otro análisis cuando las variables no se seleccionan de forma apropiada o se omiten las variables clave.

La validación es un componente clave de cualquier análisis de extracción de datos. Dada la naturaleza exploratoria de la extracción de datos y la capacidad de sus métodos de representar relaciones complejas, el investigador siempre deberá tener cuidado con que se convierta en «un análisis específico de la muestra» perdiendo, por tanto, generalización. En términos de validación interna, la mayoría de los métodos de extracción de datos utilizarán una muestra de la base de datos (incluso aunque una muestra pueda tener miles de registros). Es esencial asegurarse de que estos resultados pueden extenderse al resto de la base de datos. Aunque puede argumentarse que la validación externa no es necesaria porque a menudo todos los clientes de la empresa están representados en la base de datos, este argumento falla en última instancia a la luz de la rotación de clientes y de la continua búsqueda para adquirir nuevos clientes.

Finalmente, la evaluación de la significación práctica de los resultados es crítica por varias razones. En la investigación multivariante, muchas veces el investigador puede sentirse atraído por las técnicas y olvidar el examen de los resultados a efectos de su aplicación a la idea inicial. Tratamos aquí este asunto porque la novedad del campo y el uso de técnicas más «exóticas» consti-

tuyen una buena oportunidad para caer en este problema. En cualquier investigación, uno siempre debe examinar críticamente los resultados y determinar la auténtica medida en la que tratan la cuestión a investigar original.

Exploración frente a confirmación

Aunque las anteriores discusiones de las técnicas multivariantes se centran generalmente en aplicaciones confirmatorias, la extracción de datos tiene una naturaleza generalmente exploratoria. Este interés exploratorio no excluye una orientación a la confirmación o verificación, aunque su reducido papel se puede atribuir a varios factores. En primer lugar, por su propia naturaleza, la extracción de datos implica la búsqueda de información oculta. La cantidad de información disponible y la naturaleza incierta de muchas cuestiones clásicas de la investigación llevan en cualquier caso a un modo exploratorio con un amplio rango de relaciones posibles a explorar. También, muchas de las técnicas se diseñan para explorar una sección cruzada de opciones y a continuación profundizar en las que han dado mejores resultados. Como tales, el investigador debe gestionar el proceso exploratorio no tanto en términos metodológicos, sino para asegurar que los objetivos de la investigación se mantienen en el centro de nuestro interés.

Pero la verificación o confirmación no está «muerta». Todavía juega un papel esencial al ofrecer un apoyo directo y objetivo al descubrimiento de relaciones. La confirmación puede proceder del uso de procedimientos de validación o del uso de técnicas multivariantes más estructuradas. Como ya se ha discutido, la validación es un paso esencial porque muchas técnicas de extracción de datos se pueden ver fácilmente «sobreentrenadas» y/o perder generalidad. La validación se lleva normalmente a cabo utilizando muestras aisladas e independientes, tanto de naturaleza interna como externa. En este sentido, la validación es una simple forma de confirmación. El enfoque más riguroso es la aplicación de un modelo especificado —muchas veces una de las técnicas multivariantes discutidas en los capítulos anteriores— para replicar y confirmar los resultados. Por ejemplo, las redes neuronales pueden desarrollar tareas bastante similares a la regresión múltiple, análisis discriminante y análisis *cluster*. Muchas veces estos resultados de un procedimiento de red neuronal pueden ser reestimados utilizando la correspondiente técnica multivariante. Aunque la correspondencia directa puede ser difícil de obtener; esto ofrece al investigador un medio más estructurado y más controlable de cuantificar las relaciones descubiertas.

Técnicas de extracción de datos

El campo de la extracción de datos se acompaña de un amplio número de técnicas analíticas, que van de las más simples a las más complejas. Una característica común a todas, sin embargo, es que la mayoría de las técnicas se basa en principios muy simples que son comprensibles incluso para aquellas personas sin conocimientos estadísticos o matemáticos.

Aunque nuestras discusiones previas de las técnicas multivariantes nunca se han centrado exclusivamente en bases estadísticas, cualquier usuario debería apreciar la compleja teoría estadísticas que permite la estimación de estos modelos. Así, los textos que se centran más en la aplicación son un complemento útil para las discusiones basadas en términos matemáticos y estadísticos. Incluso hacerse con los principios básicos de las técnicas de correlación y regresión requiere cierto conocimiento. En la extracción de datos, muchas de las técnicas toman principios básicos tales como ajustarse (reglas de asociación) o analogías biológicas (redes neuronales o algoritmos genéticos) como las bases del procedimiento. La mayoría de la gente puede relacionar más rápidamente estos principios que aquellos de naturaleza estadística. Así, el investigador y los usuarios de los resultados pueden sentirse mucho mejor con los procedimientos y los resultados. Debemos prevenir, sin embargo, contra cualquier noción que en su simplicidad les haga menos «cuantitativos» o rigurosos, en la medida en que cada enfoque puede desarrollar análisis bastante específicos y detallados.

Petición* La primera técnica utilizada en muchos análisis de extracción de datos es cierta forma de **petición**. Las peticiones se consideran una técnica aparte porque su uso puede ser muy específico y en muchas ocasiones constituye el primer paso de la exploración. Una forma de petición en el almacén de datos es una petición estructurada, muchas veces conocida como SQL. Esta forma de proceso de búsqueda permite al investigador formular cuestiones muy específicas, como, «¿cuántos compradores del producto A se encuentran en los tramos de rentas medias con niños pequeños en casa y han realizado compras del producto B?». Al hacer solicitudes repetidas, el investigador puede empezar a ver o identificar pautas y relaciones. Asimismo, las peticiones permiten al investigador una búsqueda concreta, por lo que una vez que se ha identificado un hecho interesante en los datos agregados, los datos subyacentes y las pautas de identificación se exploran con más detalle.

Otra forma más estructurada de petición es el procesamiento analítico «on line» (OLAP). Esta técnica se especializa en el examen de **datos multidimensionales** (datos con dimensiones múltiples o atributos que se organizan de una forma especial en la base de datos). Esta estructura de datos facilita el examen de las intersecciones (combinaciones) de atributos y la identificación de pautas o relaciones. La naturaleza interactiva también permite la exploración de un amplio rango de relaciones posibles y a continuación una búsqueda concreta cuando se produzcan hallazgos interesantes.

Visualización Las técnicas de **visualización de datos** juegan un papel importante en la extracción de datos, permitiendo al investigador emplear uno de nuestros activos más valiosos —la capacidad de la mente humana de procesar y reconocer pautas. Tanto si se trata de simples gráficos como de enfoques más sofisticados que crean gráficos multidimensionales, el investigador puede obtener un buen conocimiento de las relaciones básicas a través de estos procedimientos. Programas tales como DIAMOND, desarrollado originariamente por IBM, se ocupa de forma específica de estas técnicas de visualización. Por ejemplo, DIAMOND ha desarrollado procedimientos de presentación de retratos multidimensionales de asociación, parecidos a como se ven en un gráfico, pero en lugar de representar en sólo dos dimensiones, puede representar hasta nueve dimensiones. El objetivo es representar relaciones de orden superior de la manera más probable para identificar la información oculta que de otra forma se vería oscurecida por los resultados más cuantitativos.

Herramientas estadísticas multivariantes Muchas de las técnicas multivariantes que ya hemos discutido se emplean en la extracción de datos, como evidencia el fuerte desarrollo de empresas como SPSS y SAS en el campo de la extracción de datos. Por ejemplo, todas las técnicas que se exponen a continuación se emplean extensivamente en la extracción de datos.

La regresión múltiple ofrece un medio directo de confirmación y algunos tipos de exploración de relaciones de dependencia. El análisis factorial se emplea en la valoración de pautas de variables y el análisis *cluster* evalúa objetivamente la asociación de objetos entre muchas variables. El análisis discriminante y la regresión logística se emplean frecuentemente a efectos de clasificación. Incluso el MDS se utiliza para presentar asociaciones conjuntas y genera adelantos de la investigación en ciertas instancias. Consecuentemente, el investigador multivariante encontrará un uso extensivo de técnicas familiares en la extracción de datos, residiendo la principal diferencia en la orientación de la investigación, como ya se ha expuesto.

Reglas de asociación Las **reglas de asociación** se parecen mucho al procedimiento de emparejamientos binarios del análisis cluster en los que se cuantifica la ocurrencia conjunta de dos eventos. Por ejemplo, considérense estas cuestiones concernientes a dos eventos: (1) ¿con qué

* o «query», es decir, filtro

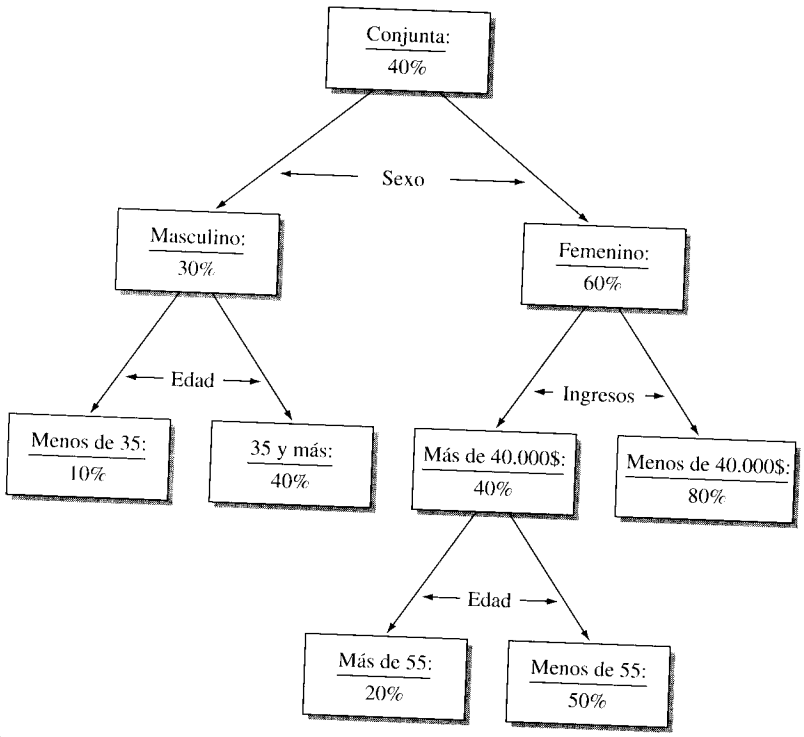
frecuencia compra cepillos de dientes cuando compra champú? y (2) ¿con qué frecuencia compra cepillos de dientes cuando compra dentífrico? Cada uno de estos casos es la base de una regla de asociación, calculada como la ocurrencia conjunta de dos eventos (el porcentaje de veces que se producen). Una regla de asociación no se restringe sólo a dos eventos, ya que podría fácilmente calcularse con qué frecuencia se compran los tres productos (pasta de dientes, cepillos de dientes y champú). Obviamente, sería simple calcular miles de reglas si el número de eventos (compras de productos en este ejemplo) se hace más grande.

¿Pero cómo podemos evaluar estas asociaciones? Se utilizan normalmente dos medidas. La primera es la **confianza**, medida como la verosimilitud de que se produzca un evento A cuando se produce un evento B. Por ejemplo, la asociación entre las compras de la marca de pasta de dientes X y los cepillos de dientes. Encontramos que un consumidor que compra una marca X de cepillos de dientes tiene una posibilidad de un 80 por ciento de comprar también una marca de pasta de dientes X. Por tanto, tenemos un intervalo de confianza del 80 por ciento de que un cliente que compra una marca de pasta de dientes X también comprará la marca de cepillos X. Además, debemos examinar también con qué frecuencia se produce este fenómeno. Denominado como **apoyo**, este es el porcentaje de tiempo en que se produce el evento conjunto sobre el total de la población. Si encontramos que sólo el 5 por ciento de los clientes compra la marca de pasta de dientes X, así como de cepillos de dientes, entonces probablemente no consideraremos esta asociación muy importante.

Las reglas de asociación son una herramienta valiosa para ajustar objetos, particularmente en un contexto de mercado, para el cual los objetos son consumidores. De nuestro ejemplo anterior, podríamos afirmar que el 5 por ciento de la muestra compraba la marca X de pasta y cepillos de dientes conjuntamente, un 80 por ciento de las veces. El investigador puede fijar a continuación niveles aceptables de apoyo y confianza con el fin de identificar grupos para una consideración ulterior.

Árboles de decisión El siguiente método, los árboles de decisión, tiene aspecto familiar (como el dendrograma del análisis cluster), pero se construye y se interpreta de forma completamente distinta. Los **árboles de decisión** son particiones secuenciales del conjunto de datos para maximizar las diferencias de la variable dependiente. Los dos programas más utilizados son CHAID (detector de interacción chi-cuadrado) y CART (árboles de clasificación y regresión). Veamos el ejemplo de partición entre compradores y no compradores mediante tres variables categóricas independientes: edad, sexo y renta (mostrado en la Figura 12.1). Dado que comprador-no comprador es dicotómico, el objetivo es identificar cuál de las tres variables ofrece la mejor escisión de compradores frente a no compradores (dado que la variable independiente es dicotómica, consideraremos sólo el porcentaje de compradores). En conjunto, el 40 por ciento de la muestra compra el producto. La mejor escisión del conjunto del grupo es con el sexo, con un 30 por ciento de hombres que compran el producto y un 60 por ciento de mujeres que compran el producto. A continuación, el procedimiento toma cada uno de estos grupos (hombres y mujeres) y encuentra la variable que mejor escinde cada grupo. Nótese que la misma variable no tiene que escindir cada grupo, en la medida en que una variable podría ser utilizada por hombres y otra por mujeres. Por ejemplo, los hombres se escinden por edad, mientras que las mujeres lo hacen por la renta. El procedimiento continúa hasta que no existen variables independientes sueltas o no existen escisiones significativas pendientes que hacer. El resultado final es un conjunto de grupos mutuamente excluyentes de clientes que varían en su porcentaje de compra.

Los resultados de nuestro ejemplo muestran que el mayor porcentaje de compradores se encuentra en mujeres con una renta por debajo de 40.000$. El grupo de menor porcentaje de compradores son los hombres menores de 35. Deberíamos anotar que los grupos que van en segundo lugar en porcentaje bajo y alto respectivamente son ambos del grupo de las mujeres, mientras que las escisiones identificadas ulteriormente identifican un grupo de mujeres con porcentaje más bajo que uno de los grupos de hombres. Cada grupo puede ser mejorado mediante la lectura del diagrama de árbol. Los árboles de decisión ofrecen una forma concisa de desarrollar grupos que son consistentes en sus atributos pero que varían en términos de la variable dependiente.



Leyenda: El valor superior de la caja es la etiqueta del grupo, el valor inferior es el porcentaje de compradores. La variable por debajo de un grupo es la utilizada para desglosar ese grupo. Por ejemplo, el sexo se utilizó para desglosar el grupo conjunto.

FIGURA 12.1. Un modelo de árbol de decisión.

Redes neuronales Las redes neuronales son unas de las herramientas más asociadas con la extracción de datos. Diseñadas después de los trabajos del sistema neuronal del cerebro, las **redes neuronales** intentan «aprender» mediante ensayos repetidos como organizarse mejor a sí mismas para conseguir maximizar la predicción. Discutiremos este método con más detalle posteriormente, pero examinemos ahora su operativa básica. El modelo se compone de **nodos**, que actúan como inputs, outputs o procesadores intermedios. Cada nodo conecta con el siguiente conjunto de nodos mediante una serie de trayectorias ponderadas (parecido a las ponderaciones en un modelo de regresión). Basado en un paradigma de aprendizaje, el modelo toma el primer caso, introduce sus datos y a continuación toma una decisión inicial basada en las ponderaciones. Se evalúa el error de predicción y a continuación el modelo hace lo mejor que puede una modificación de las ponderaciones para mejorar la predicción para seguir con el siguiente caso. Este ciclo se repite para cada caso en lo que se denomina una fase de **preparación**, cuando se esté calibrando el modelo. Una vez calibrado, el modelo se puede utilizar con otra muestra para evaluar su validez externa.

Muchos investigadores ven las redes neuronales como una «caja negra» en la que el investigador no controla la estructura del modelo, esto es, cuyos nodos o trayectorias de nodos se conectan por sí mismos. En realidad, todo esto se produce durante el proceso de aprendizaje. Pero esta estructura ofrece una gran flexibilidad. El sistema de red neuronal puede representar relaciones muy complejas, incluso no lineales —algo que es muy difícil de hacer con la mayoría de los modelos multivariantes. Además, en muchos casos puede conseguir una mayor precisión predictiva que el método estadístico comparable. Sin embargo, el investigador debe ser precavido en cuanto a la selección de muestras muy específicas y la pérdida de generalidad. Las redes neuronales se han ganado un uso más extendido en las áreas aplicadas que en las académicas, ya que aunque obtengan muy buenos resultados predictivos, lo que es necesario en las aplicaciones, se quedan cortas en las áreas académicas necesitadas de explicación. La continua investigación debería animar a ambas comunidades a apreciar más las redes neuronales.

Algoritmos genéticos El último tipo de técnicas de extracción de datos se denomina también modelo de aprendizaje, pero se basa en una analogía biológica distinta. Los **algoritmos genéticos** copian el proceso evolutivo mediante el uso de la selección natural. Empezamos con un número de soluciones posibles al problema. Las «supervivientes» de esta primera «generación» forman una nueva generación. Algunas tendrán más éxito que la generación anterior, y algunas tendrán menos éxito. Los supervivientes en cada generación sucesiva pasan y compiten por la supervivencia en la siguiente generación. Lentamente, a lo largo del tiempo, la selección natural eliminará a las soluciones peores y producirá una mejora global. Este proceso continúa hasta que se consigan unas tasas aceptables de predicción. Al automatizar este proceso, se deben resolver dos asuntos. En primer lugar, debe haber una forma de determinar a los «supervivientes». Existen muchos modos de medir el éxito, pero todos ellos están basados en alguna forma de variable dependiente. En segundo lugar, debe existir un modo para los supervivientes de formar la siguiente generación. Esto se puede conseguir mediante alguna combinación de supervivientes (utilizando una función cruzada) o mutación (una variación aleatoria para cada superviviente).

Una ventaja de los algoritmos genéticos es que generalmente convergen en la solución definitiva. Las desventajas son que esto puede llevar muchas generaciones y se requiere un gran número de individuos. Consecuentemente, este método no es particularmente eficiente, pero los avances en la potencia de cálculo han hecho posible su aplicación en un amplio rango de aplicaciones.

Una perspectiva del investigador multivariante

A lo largo de nuestra discusión, el investigador multivariante puede sentirse apartado del proceso. No se ha puesto todo el énfasis en la inferencia estadística, sino en un conjunto de modelos alternativos de enfoques diversos. E incluso cuando existe una base estadística (es decir, reglas de asociación o árboles de decisión), la significación estadística que buscamos en otros métodos multivariantes parece ser secundaria. ¿Significa esto que la estadística tiene poco que ofrecer en la extracción de datos? ¿Es un nuevo «campo» de análisis de datos que se forma con características completamente únicas?

La respuesta contundente es negativa a ambas cuestiones. Los investigadores multivariantes todavía jugarán un papel clave, pero deben entender lo que pueden añadir al proceso de extracción de datos. Sus contribuciones no se producirán a través de modelos más elaborados o sofisticados, sino que consisten en fomentar al menos cinco principios estadísticos básicos [23]. En primer lugar, retener una claridad de objetivos respecto al objeto de investigación. Cada método de extracción de datos debería ser evaluado al menos tan críticamente como cualquier técnica multivariante para su apropiada aplicación. Esto es particularmente relevante cuando se elige entre exploración, explicación y predicción. En segundo lugar, hemos de asegurarnos de que la técnica escogida es fiable en el contexto de investigación. ¿Actuará el método como se espera dado el tamaño de muestra, el número de variables y los objetivos de investigación? En tercer lugar, recordemos que todavía existe la incertidumbre. Incluso aunque la inferencia estadística del parámetro estimado no se vea resaltada, ¿qué ocurre con la probabilidad asociada con el modelo de búsqueda? Si la extracción de datos ha contrastado literalmente cientos o miles de modelos, ¿qué fiabilidad podemos dar al modelo «óptimo» final? ¿Podría producirse por azar? ¿Qué pasa con el segundo mejor modelo? ¿Ofrece los mismos resultados e interpretaciones? En cuarto lugar, hay que ser cauteloso con los datos no aleatorios. Incluso aunque el almacenamiento de datos o la base de datos pueda mantener conjuntos de datos muy grandes o incluso analizarlos, puede existir todavía cierto sesgo no aleatorio. Por ejemplo, ¿es válido el modelo para no clientes si los resultados vienen sólo de una muestra de clientes? Incluso si los resultados se van a utilizar internamente, la generalidad es esencial. Finalmente, dado que los datos son retrospectivos y no experimentales, las afirmaciones causales son muy difíciles. Se debe evitar la tentación de asignar causalidad sólo sobre la base de las técnicas utilizadas o la afirmación de causalidad. Los requisitos de las afirmaciones causales se trataron en el Capítulo 11.

Resumen

Los campos del almacenamiento y la extracción de datos son un contexto de investigación muy distinto al que se encuentra normalmente un investigador. Pero estas diferencias no deberían excluir contribuciones de ambas áreas, y quizá se pueda producir una «simbiosis de estadísticas» a la que cada campo contribuya igualmente [23]. La colaboración ayudará a cada parte a desarrollar el campo en su totalidad, así como sus intereses particulares. En una perspectiva más amplia, esta interacción podría ofrecer un prototipo de tal cooperación en áreas con distintas constituciones conceptuales y aplicadas.

Redes neuronales

Las redes neuronales son un procedimiento totalmente diferente de analizar los datos respecto a cualquiera de las técnicas multivariantes. En lugar de conceptualizar el problema como matemático, las redes neuronales utilizan el cerebro humano y su estructura para desarrollar una estrategia de procedimiento. Aunque nunca seremos capaces de construir redes neuronales tan complejas como el cerebro humano, podemos utilizar sus principios básicos de unidades de procesamiento múltiple involucradas en el reconocimiento de pautas. Las redes neuronales difieren no sólo en su estructura, sino también en proceso. Un elemento clave de una red neuronal es el aprendizaje (otra analogía con el cerebro humano), mediante el que los errores de los resultados (predicción o clasificación) se retroalimentan en el sistema y se ajustan consecuentemente. A continuación el proceso se repite, aprendiendo de cada conjunto de errores de resultado. Se trata de un proceso secuencial, una vez por instante, en comparación con otra técnica multivariante, que considera el conjunto completo de casos simultáneamente.

Las redes neuronales se hicieron operativas por primera vez en los últimos años de la década de los 50. Sin embargo, a finales de los 60, la investigación demostró que las redes neuronales de aquel momento eran bastante limitadas en capacidad, y el proceso de investigación sufrió un estancamiento. El interés volvió en los 80 a medida que al desarrollo teórico se unió un mayor poder de cálculo. La principal mejora fueron las capas ocultas (se expondrán más detalles en una sección ulterior), que permitieron a las redes neuronales representar sistemas mucho más complejos. Hoy en día, las redes neuronales se utilizan en casi todas las disciplinas o áreas de análisis. La naturaleza flexible de las especificaciones del sistema las hace adaptables a un amplio rango de problemas, que van desde la predicción a la clasificación o incluso al análisis de las series temporales.

En primer lugar, clarifiquemos las relaciones entre las redes neuronales y las técnicas estadísticas multivariantes. Las redes neuronales pueden tratar muchos de los problemas que tratan las técnicas multivariantes de la regresión múltiple, el análisis discriminante y el análisis de conglomerados. En muchos casos, las redes neuronales producen resultados comparables [1, 24], y por tanto depende del investigador seleccionar entre los métodos de acuerdo a los objetivos de investigación. Asimismo, las redes neuronales tienen una base estadística subyacente, por ejemplo, el impacto de las distribuciones de *inputs* (no normales) sobre la estimación de las ponderaciones. La principal diferencia de las técnicas multivariantes es la ausencia de contrastes de inferencia estadística para las ponderaciones del ajuste global del modelo. Sin embargo, el investigador no debe tomar las redes neuronales como algo menos riguroso, sino más bien como una variación del enfoque.

La discusión de este capítulo arroja sólo una breve introducción al campo de las redes neuronales. El interés del investigador se refiere a cualquier número de introducciones que utilicen las redes neuronales en un amplio rango de aplicaciones [5, 7, 11, 15, 16, 29, 35]. Nuestra discusión se centra en primer lugar en algunos conceptos básicos de las redes neuronales y a continuación explora los pasos necesarios para estimar una red neuronal. Finalmente, se presenta un breve ejemplo comparando un modelo de red neuronal con los resultados del análisis discriminante del Capítulo 5.

Conceptos básicos de las redes neuronales

Las redes neuronales tienen una estructura simple y operativa que puede describirse mediante cuatro conceptos: (1) el tipo de modelo de red neuronal; (2) las unidades de procesamiento (nodos*) que recogen información, procesan dicha información y crean un valor resultante; (3) el sistema de nodos (red) organizado para transmitir señales desde los nodos de entrada a los nodos de salida, con algunos nodos intermedios entre ellos; y (4) la función de aprendizaje mediante la cual el sistema «retroalimenta» los errores de predicción para recalibrar el modelo. Discutimos cada uno de estos conceptos y cómo se interrelacionan con más detalle en la siguiente sección.

Tipos de modelos de redes neuronales

Existen tres tipos de redes neuronales: el perceptor multicapa, la función de base radial y las redes de Kohonen. El modelo del perceptor multicapa es el más común y es el tipo que se utiliza en el ejemplo que se expondrá a continuación. La **función de base radial** es un método desarrollado más recientemente que se puede utilizar para las mismas tareas que el modelo multicapa, pero que trabaja de una manera diferente. El **modelo de Kohonen** es apropiado sólo para los problemas de conglomerados. El lector interesado puede encontrar fácilmente discusiones de la fuerza y debilidad de cada red en diversas fuentes [5, 11, 15, 29, 35]. Todas nuestras discusiones se centran en el modelo de perceptor multicapa.

Nodos

El elemento más básico de una red neuronal es un **nodo**, una unidad de procesamiento autocontenido que actúa en paralelo con otros nodos de una red neuronal. El nodo es análogo a la neurona del cerebro humano, que acepta *inputs* y genera *outputs*. Una representación simple del nodo y su forma de actuar se muestra en la Figura 12.2, que muestra que los nodos aceptan un número de *inputs* de otras fuentes (nodos). Cada conexión de otro nodo tiene un valor asignado. La primera tarea del nodo es procesar los datos de entrada creando un valor resumen en el que cada valor de entrada se multiplique por su respectiva ponderación. (Nótese que ésta es la misma operación desarrollada por nuestros métodos multivariantes cuando se calcula el valor teórico.) Este valor resumen, se procesa a continuación mediante una **función de activación** para generar un valor de salida, que se envía al siguiente nodo del sistema. Las funciones de activación son generalmente funciones no lineales, como la **función sigmoide**, que es una clase general de curvas en forma de S que incluyen la función logística (véase Capítulo 5 para una discusión más detallada de las funciones logísticas).

Red neuronal

La red neuronal es una ordenación secuencial de tres tipos básicos de nodos o capas: nodos de entrada, nodos de salida y nodos intermedios (escondidos) (véase Figura 12.2). Los nodos de entrada reciben valores iniciales de los datos de cada caso y los transmiten a la red neuronal. Un nodo de entrada representa una única variable o pauta. Las variables métricas sólo requieren un nodo por variable. Las variables no métricas tienen que estar **codificadas**, lo que significa que cada categoría se representa por una variable binaria (igual que la creación de variables ficticias binarias del Capítulo 2, sólo que ninguna de ellas se destruyen como categoría de referencia). Por tanto, se pueden representar tres variables no métricas mediante tres nodos de entrada

* En algunos casos, se ha traducido nodo (*node* en inglés) por nudo. Preferimos conservar la denominación que más se ajusta a la expresión original.

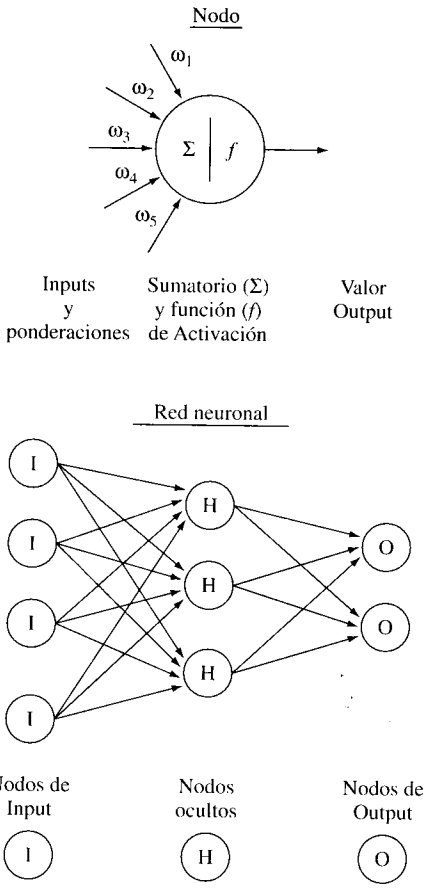


FIGURA 12.2. Nodos y red neuronal.

binarios. La primera categoría tendría valores de 1, 0, 0 para las tres variables, la segunda categoría tendría valores de 0, 1, 0 y la tercera tendría valores de 0, 0, 1. Un nodo de salida recibe entradas y calcula un valor de salida, pero en lugar de ir a otro nodo, éste es el valor final. Si se trata de un modelo predictivo, entonces éste es el valor predictivo. Si el modelo se utiliza para la clasificación, entonces éste es el valor utilizado en los procesos de clasificación (el lector puede revisar si lo desea el Capítulo 5 en relación con los procesos de clasificación del análisis discriminante).

En casi todas las redes neuronales, existe un tercer tipo de nodo contenido en la **capa oculta**. Es un conjunto de nodos utilizados por la red neuronal para representar relaciones más complejas que las relaciones uno a uno de *input* a *output*. Es la capa oculta y la función de activación la que permite a las redes neuronales representar fácilmente las relaciones no lineales, que son muy problemáticas para las técnicas multivariantes. Este diseño de redes neuronales permite a cada nodo actuar independientemente, pero en paralelo, con el resto de nodos. Por ello, ofrece a la red neuronal una gran flexibilidad en los tipos de relaciones *input-ouput* que puede manejar. En una sección posterior, discutiremos diversos asuntos en relación con el número de nodos y las capas ocultas.

Aprendizaje

La característica de una red neuronal, que verdaderamente la aparta del resto de las técnicas multivariantes, es su capacidad para «aprender» o «corregirse a sí misma» basándose en sus errores. Previamente, discutimos los conceptos de ponderaciones en las conexiones entre nodos. Utilizando una analogía biológica, las ponderaciones representan un estado de memoria, el «me-

jor cálculo» de cómo hacer predicciones a partir de los resultados de los nodos. Una vez que la entrada de un caso se procesa a través del sistema, puede compararse con el valor del resultado efectivo. A esto se le denomina preparación del sistema en un modelo de aprendizaje **supervisado**. Los valores resultantes y los efectivos se comparan. Si existe alguna diferencia entre los dos valores (parecido a un valor residual), entonces nos gustaría ajustar el modelo con la esperanza de mejorarlo. La forma más común de preparación es la **retropropagación**. En esta aproximación, se calcula el error en el valor del ouput y se distribuye hacia atrás en el sistema. Como funciona por su vía a través del sistema de nodos, las ponderaciones cambian proporcionalmente, aumentando o disminuyendo dependiendo de la dirección del error. Una vez que todas las ponderaciones se han recalibrado, se introduce en la red el *input* de otro caso y se comienza de nuevo el proceso. El objetivo es procesar un gran número de casos a través de la red neuronal en la fase de preparación, de tal forma que pueda hacer las mejores predicciones para todas las pautas de entrada de datos. Existe también un modo sin supervisar en el que no se da ninguna retroalimentación hacia lo que es el valor correcto de *output*. Esta aproximación se utiliza sólo en los problemas de *clustering* porque no hay forma de conocer las soluciones efectivas del conglomerado.

Estimación de un modelo de red neuronal

Ahora que hemos descrito las operaciones básicas y los componentes de la red neuronal, nos centramos en cinco aspectos fundamentales implicados en la estimación efectiva de un modelo de red neuronal: preparación de los datos, definición de la estructura del modelo, estimación del modelo, evaluación de los resultados del modelo y validación del modelo.

Preparación de los datos

Las redes neuronales son como el resto de los métodos estadísticos en un sentido: si metes basura, sacarás basura. Las redes neuronales no tienen la capacidad de transformar la baja calidad o el mal acondicionamiento de los datos en un buen modelo. El investigador, por tanto, debe examinar los datos junto con cualquier otro método estadístico. Al preparar los datos para su uso en una red neuronal, el investigador debe considerar: (1) el tamaño muestral, y (2) las distribuciones, transformaciones y codificaciones de los datos.

Tamaño muestral La primera tarea es crear las bases de datos para la estimación de su red neuronal. Se necesita una muestra de calibración, conocida como muestra de preparación, utilizada para estimar las ponderaciones y una muestra distinta de validación para evaluar independientemente la capacidad del modelo. La escisión de la muestra original está basada primordialmente en el tamaño de muestra exigido para las muestras de calibración y validación.

El tamaño muestral utilizado para calibrar un modelo de red neuronal puede tener tanto impacto sobre los resultados como con cualquier otra técnica multivariante. En primer lugar, un modelo puede estar perfectamente ajustado si el tamaño muestral es menor que el número de parámetros estimados. Pero a medida que el número de casos se aproxima al número de coeficientes, se produce el sobreajuste y el modelo se hace muy específico para la muestra, perdiendo generalidad [30]. El número de ponderaciones en un modelo simple de red neuronal puede aumentar rápidamente. El modelo de red neuronal de la Figura 12.2 tiene 12 coeficientes entre los *inputs* y la capa oculta y otros seis coeficientes entre la capa oculta y los nodos de resultado por un total de 18 ponderaciones. Por tanto, el número de ponderaciones se relaciona con el número de capas y el número de nodos en cada capa. Añadiendo cada vez un nodo de entrada se añadirían tres coeficientes, mientras que añadiendo una segunda capa oculta de tres nodos se añadirían nueve coeficientes.

El modelo requiere que haya algo más que un caso adicional al número de coeficientes. La norma habitual es tener entre 10 y 30 casos en la base de datos de calibración para cada parámetro

estimado. Aunque esto no es más que una recomendación, debería considerarse seriamente cuando se diseñe la red. Habría que decir que nunca se tienen realmente muchos casos (al contrario que en los test estadísticos, en los que la potencia del test aumenta notablemente y los contrastes de inferencia estadística siempre se muestran significativos). Los casos se pueden muestrear aleatoriamente si el tamaño de muestra original se juzga que es muy grande para procesarlo o resulta inabarcable para el investigador.

Examen de los datos Incluso aunque no se utilicen los test de inferencia estadística, el investigador debería seguir examinando los datos con las técnicas presentadas en el Capítulo 2 con el fin de evaluar la simetría, la no normalidad y los casos atípicos. Respecto a las variables métricas, existen técnicas para evaluar estas características de una forma fácil y llevar a cabo las transformaciones si fuera necesario. Las variables no métricas pueden resultar problemáticas, particularmente si el nodo de *output* es una variable métrica simétrica. Si existen dos categorías, no es bueno tener el 98 por ciento de los casos en una categoría y el 2 por ciento en la otra. El modelo de red neuronal se concentrará en el valor grande (98 por ciento) porque el error producido por la pérdida de la otra categoría es reducido. El investigador debería buscar un tamaño parecido para cada categoría. Si esto no se produce de forma natural, se anima al investigador a igualar las categorías a través de la muestra. El problema es que el tamaño total de la muestra está limitado por el grupo de *output* más pequeño.

Además de la simetría y la normalidad, los datos también deberían examinarse en busca de puntos perdidos, siguiendo las soluciones recomendadas en el Capítulo 2. Los casos atípicos deberían examinarse cuidadosamente para su eliminación, dado que se pueden ver seriamente afectados por los resultados conjuntos y los procedimientos de preparación. El último aspecto se centra en la estandarización de las variables. La escala utilizada para medir las variables puede provocar la estimación de grandes coeficientes para los valores más altos de la escala. Aunque existen muchas sugerencias sobre cuándo estandarizar, este proceso no tiene un riesgo real, y sugerimos que todas las variables se estandaricen. Una última nota, es posible añadir algo de ***jitter*** (ruido blanco) a la entrada de datos, para facilitar el proceso de estimación y suavizar el ajuste de la función de tal forma que tenga una mejor oportunidad de alcanzar una solución óptima global (véase una sección ulterior para una discusión más detallada).

Definición de la estructura del modelo

Una vez que se han examinado los datos y se han realizado las transformaciones que sean pertinentes, el investigador debe especificar la estructura del modelo. Dado que los *inputs* y los *outputs* ya se han seleccionado, en este paso se deciden el número de capas ocultas y el número de nodos en cada capa. Una capa oculta adicional puede mejorar ligeramente la estimación, pero con un riesgo mayor de sobreajuste y de encontrar sólo una solución subóptima, aumentando también el tiempo necesario para la estimación. Si se añade una segunda capa, se pueden utilizar procedimientos de validación de forma rigurosa. En lo que se refiere al número de nodos en una capa oculta, se trata de un método de ensayo y error. El investigador sólo debería modificar el número de nodos en la capa oculta para encontrar el mejor ajuste del modelo. Si las dos soluciones tienen igual ajuste, coja la estructura más simple (menor número de nodos). Un último elemento es el **nodo sesgado**, que es una clase de nodo oculto pero que tiene un valor constante de 1,0. Se puede añadir para que actúe como un término constante en una ecuación de regresión.

Estimación del modelo

El principal objetivo es conseguir el mejor ajuste posible del modelo buscando la solución global óptima y no la sobrepreparación. La solución global óptima es la mejor solución posible de todas las soluciones posibles, denominadas el **espacio de búsqueda**. Dado que el proceso de

estimación es un procedimiento iterativo y que busca modos de ajuste incrementales, es posible que no siempre se encuentre la solución óptima. Por ejemplo, la Figura 12.3 representa una solución posible subóptima y una solución óptima global. Si el punto de partida estuviera más a la derecha, la solución puede mejorarse a medida que se mueve hacia la izquierda, pero cuando se alcanza el fondo del «valle», no podría hacer ninguna mejora. El proceso no tiene ninguna forma de «saltar» el valle, de tal forma que puede proceder hasta el óptimo global. Sin embargo, si el punto de partida estuviera a la izquierda, se alcanzaría el óptimo global. Esto ilustra la necesidad de utilizar puntos de partida múltiples para asegurar que ha alcanzado el óptimo global.

Sabemos que todos nuestros datos tienen algún error o ruido aleatorio, así que no esperemos un ajuste perfecto (al igual que no esperaríamos un ajuste perfecto en nuestros modelos multivariantes). Pero si un modelo se entrena con muchos casos, es posible que se sobreentrene, lo que significa que ha explicado todas las relaciones básicas entre los nodos y que comienza a representar el error aleatorio del modelo. El investigador puede evitar este problema de dos formas. La primera es fijar un límite inferior de error y detener el entrenamiento cuando se alcance. Aunque sea arbitrario, evita directamente el problema. Un segundo enfoque, es controlar la tasa de error tanto para las muestras de calibración como de validación. Cuando la red neuronal se está preparando inicialmente, el ajuste mejorará ambas muestras así como las ponderaciones se calibran. Pero en un punto, el error de la muestra de validación se nivelará o incluso empezará a aumentar, divergiendo de la muestra de calibración. Éste es el punto en el cual la muestra de calibración se empieza a sobrepreparar, es decir, se hace muy específica para la muestra (y no generalizable). Esto también produce un bajo ajuste de la muestra de validación. El investigador debería detenerse cuando se haya encontrado la mejor de las soluciones posibles.

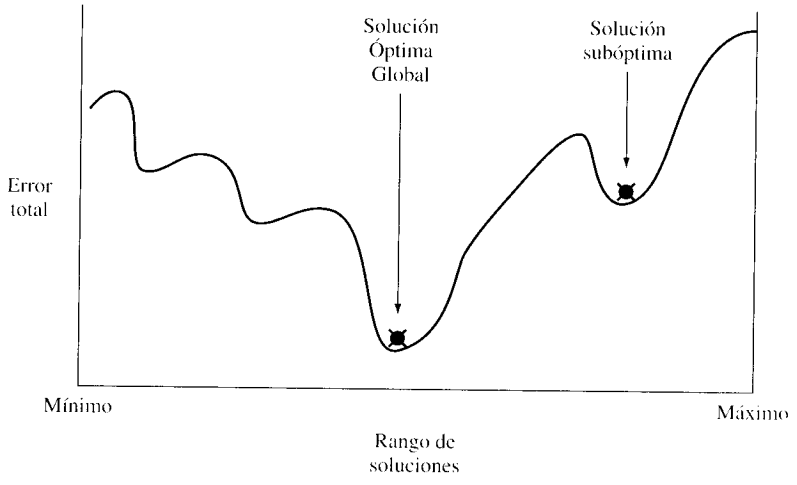


FIGURA 12.3. Óptima global y soluciones subóptimas.

Evaluación de los resultados del modelo

La evaluación de un modelo de red neuronal consiste fundamentalmente en la evaluación del nivel de predicción o clasificación de las variables de entrada. Por ejemplo, en un problema de clasificación, la matriz de clasificación tabula de forma cruzada los valores actuales y previstos situando las predicciones corregidas en la diagonal. Desde este número, se puede calcular el porcentaje correctamente clasificado y comparar con diferentes criterios (véase la discusión del Capítulo 5 sobre la precisión predictiva en el análisis discriminante con más detalle). Si el *output* es una variable métrica, entonces la medida común de ajuste es el error cuadrático medio.

Incluso aunque un modelo de red neuronal pueda desarrollar tareas comparables a las otras técnicas multivariantes, tales como la regresión múltiple o el análisis discriminante, no ofrece información sobre la importancia relativa de las variables de entrada. Aunque existen ponderaciones para

cada variable de entrada, no son directamente interpretables porque también deben combinarse con las ponderaciones de la capa oculta de la variable de salida. Hasta el momento no hay un método simple de interpretación de la solución en relación con la importancia del impacto de una única variable de entrada.

Validación del modelo

El paso final es validar la solución para asegurar que es el óptimo global y que es tan generalizable como sea posible. Como ya se ha mencionado, es esencial que se cree una muestra de validación para ofrecer una evaluación independiente del ajuste del modelo que no dependa de la muestra de calibración. Mientras sea posible, el investigador debería emplear una nueva muestra de casos para una evaluación adicional del ajuste. La estabilidad de la solución se evalúa ofreciendo diferentes puntos de partida para las ponderaciones y reorganizando el orden de los casos de la muestra de calibración. Finalmente, el investigador también debería variar el número de nodos para asegurarse de que no es posible una mejor solución.

Resumen

El proceso de estimación de un modelo de red neuronal no se lleva a cabo en un único paso, sino que requiere que el investigador evalúe constantemente las soluciones del modelo a efectos de ganar precisión predictiva, vigilar el sobreentrenamiento o el sobreajuste. El investigador debería «experimentar» con la solución en la medida en que sea posible porque el procedimiento no siempre garantiza una solución óptima global o general. Cualquier investigador que desee utilizar este procedimiento debería ser consciente de las consideraciones a cada paso, dado que no es un método en el que baste con utilizar valores por «defecto» y estar seguro de obtener un modelo aceptable.

Utilización de una red neuronal para la clasificación

Una de las aplicaciones más comunes de las redes neuronales son los problemas de clasificación: decidir a qué grupo pertenece una observación. Esto se corresponde al análisis discriminante y al problema de regresión logística. Para ofrecer una comparación entre las capacidades del análisis discriminante y las redes neuronales, en este apartado se revisará el problema de los dos grupos del análisis discriminante. Además de los resultados del análisis discriminante, también estimamos cuatro modelos de red neuronal. Los tres primeros, tienen una capa oculta con cuatro, seis y ocho nodos. El cuarto modelo de red neuronal tiene dos capas ocultas con cuatro nodos cada una. Los modelos de red neuronal alternativos se estiman para averiguar la precisión predictiva que se gana a través del aumento de la complejidad del modelo.

La Tabla 12.1 contiene las descripciones del modelo de red neuronal y los resultados predictivos. La precisión de la clasificación se mide por el porcentaje de la clasificación correcta (véase Capítulo 6 para una descripción más detallada si fuera necesario). En primer lugar, el modelo del análisis discriminante predice muy bien, con un porcentaje del 97,5 por ciento de acierto. Pero el modelo de red neuronal supera al análisis discriminante para mayores grados de complejidad del modelo. Los niveles de precisión de clasificación predictiva de los modelos de red neuronal van desde el 80 por ciento del modelo más simple al 100 por cien de los modelos más complejos. Nótese que aunque no era posible ganar en precisión predictiva para los dos modelos más elevados, la suma de la segunda capa oculta realmente aumenta la tasa de error de preparación. Esto apoya la tesis de que con una única capa oculta es suficiente en la mayoría de los casos.

Los investigadores deberían considerar la aplicación de redes neuronales para los problemas de predicción y clasificación, especialmente cuando se hace más énfasis en la precisión de la clasificación y no en la interpretación del valor teórico. Como hemos visto, las redes neuronales pue-

TABLA 12.1. Comparación de la red neuronal y los modelos de análisis discriminante: Dos grupos de clasificación

Modelo de red neuronal				
Modelo	*Estructura*[a]	*Número de ponderaciones*	*Error de preparación*[b]	*Precisión de clasificación*[c]
NN1	7-4-1	32	0,196	80,0%
NN2	7-6-1	48	0,142	90,0%
NN3	7-8-1	64	0,026	100,0%
NN4	7-4-4-1	56	0,034	100,0%
Análisis discriminante				
Modelo de dos grupos	97,5%			

[a] El primer valor es el número de nodos de entrada. El último valor es el número de nodos de salida. Los valores de en medio representan los nodos en las capas ocultas.
[b] Error cuadrático medio (RMS).
[c] Porcentaje correctamente clasificado en la muestra de contraste (red neuronal) o muestra de validación (análisis discriminante).

den ajustarse e incluso superar la capacidad predictiva de las técnicas multivariantes apropiadas. La creciente capacidad del *software*, como NEURAL CONNECTION de SPSS, facilita enormemente la adopción de esta técnica.

Resumen

Las redes neuronales presentan una aproximación al análisis de datos que es claramente diferente de la que cubre en este texto. Su capacidad para manejar relaciones complejas, particularmente aquellas de naturaleza no lineal, ofrece un instrumento analítico de gran capacidad para los tipos de problemas que pueden tratarse. Esta flexibilidad proporciona la base de una superior estimación de resultados en muchos problemas de predicción y clasificación. No ofrece, sin embargo, interpretación de la importancia relativa de las variables de entrada, ni de sus interconexiones. Por tanto, el investigador debería emplear redes neuronales en aquellas situaciones en las que (1) las técnicas multivariantes no puedan tratar las relaciones complejas, y (2) la clasificación y la predicción sean los objetivos primordiales.

Remuestreo

La clase de estadística no paramétrica que se pensó en el pasado jugó un importante papel en el análisis de datos que no eran continuos y que, por tanto, no podían emplear la distribución de probabilidad normal al hacer estimaciones del intervalo de confianza y del parámetro. Con ello, no hemos supuesto que el intervalo de confianza de un parámetro siga una distribución normal. Podemos incluso generar intervalos de confianza para parámetros tales como la mediana, que normalmente es difícil de evaluar con las técnicas de inferencia paramétrica tradicionales. Este enfoque no paramétrico, conocido generalmente como resmuestreo, ha ido ganando partidarios como alternativa a nuestros métodos de inferencia paramétrica. El **remuestreo** descarta la distribución muestral supuesta de un parámetro y calcula una distribución empírica —la distribución real del parámetro— para cientos o miles de muestras. Con el remuestreo, no tenemos que basarnos en la distribución supuesta o ser cautelosos sobre el incumplimiento de uno de los supuestos subyacentes. Podemos calcular una distribución efectiva de los parámetros de la muestras y podemos ver ahora dónde están realmente los percentiles del 95 o del 99.

Pero, ¿de dónde proceden las diversas muestras? ¿Es necesario obtener muestras separadas, aumentando notablemente el esfuerzo de recogida de datos? A lo largo de los años, los estadísticos han desarrollado varios procedimientos para crear las múltiples muestras necesarias para remuestrear *desde la muestra original.* Ahora, una muestra puede generar un gran número de otras muestras que pueden utilizarse para generar la distribución muestral empírica. En esta sección, revisaremos someramente la estimación paramétrica y a continuación discutiremos algunos de estos métodos, centrándonos principalmente en los métodos de *jackknife* o *bootstrapping**. Una vez examinados algunos de los aspectos implicados en su uso, presentaremos un ejemplo sencillo de remuestreo aplicado a un modelo de regresión múltiple.

Un breve repaso de la inferencia paramétrica

En primer lugar, examinemos exactamente lo que quiere decir la inferencia paramétrica. Primero, cuando estimamos un parámetro (por ejemplo, la media) de una muestra, sabemos que el valor calculado es sólo una estimación y que tendremos otros valores para otras muestras. A medida que las muestras se hacen más grandes, la variación de la media de la muestra se hace más pequeña, pero todavía existe cierta variación. Así que no importa qué tamaño muestral tengamos, necesitamos saber no sólo cuál es el valor del parámetro estimado, sino cuánto podemos esperar que varíe para todas las muestras posibles de este tamaño.

La teoría estadística clásica ofrece un modo sencillo de hacer estas estimaciones. Si suponemos que la media seguirá una distribución normal, lo que significa que los valores subyacentes de la muestra se distribuyen normalmente, entonces podemos calcular el error estándar de la media (donde el error estándar es la desviación estándar esperada de las medias estimadas a lo largo de todas las muestras). Por tanto, $\pm 1{,}96$ de error estándar nos darían un 95 por ciento de intervalo de confianza para nuestra media calculada. Ahora tenemos la media calculada y la medida de cómo variaría.

La estadística teórica ofrece ciertas garantías en las que el teorema central del límite muestra que las medias estarán distribuidas normalmente para grandes muestras incluso si los valores subyacentes no lo están. Pero no podemos asegurar que se mantendrá para muestras pequeñas, y los intervalos de confianza de muchos parámetros serán difíciles de calcular.

Conceptos básicos de la reespecificación

El remuestreo, sin embargo, no utiliza la distribución de probabilidad supuesta. En su lugar calcula una distribución empírica de los parámetros estimados. Mediante la creación de muestras múltiples desde la muestra original, la reespecificación ahora sólo necesita potencia de cálculo para estimar un valor de parámetro para cada muestra. Una vez que se han calculado todos, podemos examinar un histograma de los valores e incluso calcular los intervalos de confianza a partir de la distribución efectiva de los parámetros estimados. La teoría subyacente de reespecificación no se analiza en nuestra discusión actual, pero podemos encontrarla en otras fuentes [9, 10, 25]. El remuestreo utiliza en realidad métodos bien conocidos. Discutimos aquí los más utilizados: el *jackknife* y el *bootstrap.*

Métodos de remuestreo

Una diferencia clave entre los diferentes métodos de remuestreo se encuentra en si las muestras se extraen con o sin reemplazamiento. Las muestras con reemplazamiento extraen una observa-

---

* Hemos decidido no traducir estos dos términos, puesto que en la comunidad académica se utilizan normalmente estas expresiones en inglés.

ción de la muestra y a continuación la devuelven a la muestra posibilitando su muestreo ulterior. El muestreo sin reemplazamiento deduce observaciones desde la muestra, pero una vez deducidas no se vuelven a incluir para volver a ser extraídas. La potencia verdadera del remuestreo proviene del remuestreo con reemplazamiento. El investigador ha mostrado que este método ofrece la estimación directa de los intervalos de confianza, aunque existan varias modificaciones de los métodos simples para obtener intervalos de confianza. Revisemos en qué difieren el *bootstrap* y el *jackknife*.

Jackknife versus bootstrap Los métodos *bootstrap* y *jackknife* difieren en cómo obtienen la muestra. El método *jackknife* [32] calcula subconjuntos (n = tamaño muestral) mediante la eliminación secuencial de un caso en cada muestra. Por tanto, cada muestra tiene un tamaño muestral de $n - 1$ y difiere sólo en el único caso omitido en cada muestra. Aunque el método de *jackknife* ha sido sobrepasado por el *bootstrap* como estimador eficiente de los intervalos de confianza, todavía se mantiene como una medida viable de observaciones influyentes (véase Capítulo 4 para una discusión más detallada de las observaciones influyentes) y como opción en muchos paquetes estadísticos.

El método *bootstrap* deduce su muestra mediante el muestreo con reemplazamiento de la muestra original. La clave es el reemplazamiento de las observaciones después del muestreo, lo que permite al investigador crear tantas muestras como sean necesarias y no preocuparse nunca por la duplicación de las muestras a no ser que sea por azar. Cada muestra puede ser analizada independientemente y los resultados pueden recogerse para varias muestras. Por ejemplo, la mejor estimación de la media es justamente la media de todas las medias estimadas para todas las muestras. También puede calcularse directamente el intervalo de confianza. Los dos enfoques más sencillos (1) calculan el error estándar de forma tan simple como las desviaciones estándar de las medias estimadas, o (2) literalmente ordenan las medias estimadas y definen los valores que contienen el 5 por ciento (o el 1 por ciento) fuera de los valores medios estimados. Otros enfoques añaden algunos factores de corrección a este proceso, y son notablemente más complejos en sus cálculos [25].

Limitaciones

Mientras que los procedimientos de reespecificación no están limitados por ningún supuesto paramétrico, todavía tienen ciertas limitaciones [25]. En primer lugar, y quizá la más importante, la muestra debe ser lo suficientemente grande y deducida supuestamente de forma aleatoria de manera que sea representativa del conjunto de la población. Las técnicas de reespecificación no pueden obviar los sesgos debidos a una muestra no representativa y en muchos casos exacerbarían el problema. En segundo lugar, los métodos paramétricos son mejores en muchos casos para realizar estimaciones puntuales, tales como la media conjunta. Por tanto, los procedimientos de reespecificación pueden complementar las estimaciones puntuales de los métodos paramétricos ofreciendo estimaciones de los intervalos de confianza. Finalmente, las técnicas de reespecificación no son adecuadas para la identificación de parámetros que tienen un rango muestral muy reducido, tales como valores máximos o mínimos. La reespecificación funciona mejor cuando se considera la distribución entera.

Un ejemplo de reespecificación y regresión múltiple

Como ilustración de las intuiciones ofrecidas por la reespecificación, la regresión múltiple se realiza utilizando tanto los parámetros estimados como las estimaciones de *bootstrap*. El problema es similar al ejemplo utilizado en el Capítulo 4, excepto en que se desarrolla una solución simultánea (todas las variables retenidas en la solución). Se utilizan siete variables independientes (X_1 a X_7) para predecir una variable métrica dependiente (X_9). En el Capítulo 1 se encuentra una explicación más detallada de las variables.

La Tabla 12.2 contiene las estimaciones de la regresión (coeficientes y errores estándar) para cada variable. Sólo dos de las variables (X_3 y X_5) son estadísticamente significativas.

También se muestra en la Tabla 12.2 las estimaciones *bootstrap* para tamaños de muestra de 50, 100, 1.000 y 2.500. Los tamaños de muestra se refieren al número de estimaciones aisladas para cada coeficiente. Cada estimación exige una muestra aislada, para conseguir el tamaño muestral de 2.500 exigida para la generación de 2.500 muestras *bootstrap* (cada una con un tamaño de 100) y a continuación la estimación de la ecuación de regresión para cada muestra. Los errores estándar para las muestras del *bootstrap* se calcularon con el método de aproximación normal.

La comparación de los coeficientes y de los errores estándar entre las muestras de regresión y *bootstrap* indica que la evaluación de la significación estadística sería idéntica para ambos métodos. Aunque las variables no significativas muestran una clara volatilidad para todos los tamaños muestrales y entre las estimaciones de regresión y *bootstrap*, lo que es esperable porque tiene una menor correlación con las variables dependientes y/o multicolinealidad con las otras variables independientes. Las dos variables significativas estadísticamente (X_3 y X_5) son bastante similares, aunque los coeficientes para X_5 son normalmente mayores en las estimaciones de *bootstrap*.

La Figura 12.4 representa la distribución muestral de *bootstrap* para cada uno de los siete coeficientes para un tamaño muestral de 1.000. Cuatro de los coeficientes (X_3, X_4, X_6 y X_7) se aproximan bastante a la distribución normal supuesta en la estimación paramétrica. Las otras tres estimaciones, sin embargo, difieren notablemente. Tanto X_1 como X_2 tienen una asimetría a izquierdas, mientras que X_5 tiene la orientación opuesta. Además, tres distribuciones simétricas (X_1, X_6 y X_7) son mucho más apuntadas que la distribución normal. Esta simetría genera errores estándar sustancialmente elevados, a medida que el tamaño muestral aumenta para estas tres variables. Esto indicaría que las estimaciones paramétricas son más pequeñas que las efectivas y por tanto los test estadísticos son tolerantes.

La Figura 12.5 demuestra los cambios en la forma de la distribución empírica a medida que aumenta el tamaño muestral. Para pequeños tamaños muestrales la distribución está menos bien formada, pero a medida que el tamaño muestral aumenta hasta 1.000 o más, la distribución se conforma bien y en este caso se aproxima a la distribución normal. Esto sugiere que las muestras de *bootstrap* deberían basarse en al menos 1.000 muestras. Aunque esto puede parecer en primer lugar problemático, los computadores personales pueden tratar fácilmente problemas de este tamaño.

TABLA 12.2. Comparación de coeficientes de regresión y errores estándar obtenidos por estimación paramétrica y *bootstrapping*

	Estimación de regresión		*Estimaciones de bootstrap para el tamaño de muestra*				
Variable	*Coeficiente*[a]	*Significación*	50	100	500	1000	2500
X_1	−0,058	0,977	−0,050	−0,720	−0,323	−0,512	−0,527
	(2,013)		(1,820)	(2,799)	(2,351)	(2,458)	(2,409)
X_2	−0,697	0,740	−0,751	−1,491	−0,971	−1,148	−1,196
	(2,090)		(1,831)	(2,743)	(2,431)	(2,492)	(2,460)
X_3	3,368	0,000	3,425	3,287	3,374	3,376	3,352
	(0,411)		(0,406)	(0,384)	(0,376)	(0,381)	(0,384)
X_4	−0,042	0,950	−0,105	0,078	0,033	0,039	0,043
	(0,667)		(0,533)	(0,624)	(0,636)	(0,591)	(0,598)
X_5	8,369	0,035	8,535	9,916	8,996	9,301	9,372
	(3,918)		(3,255)	(5,348)	(4,567)	(4,743)	(4,677)
X_6	1,281	0,180	1,418	1,099	1,183	1,215	1,169
	(0,947)		(0,923)	(1,135)	(1,116)	(1,044)	(1,089)
X_7	0,567	0,114	0,664	0,599	0,558	0,581	0,587
	(0,355)		(0,363)	(0,301)	(0,322)	(0,365)	(0,347)

[a] Los errores estándar se muestran entre paréntesis.

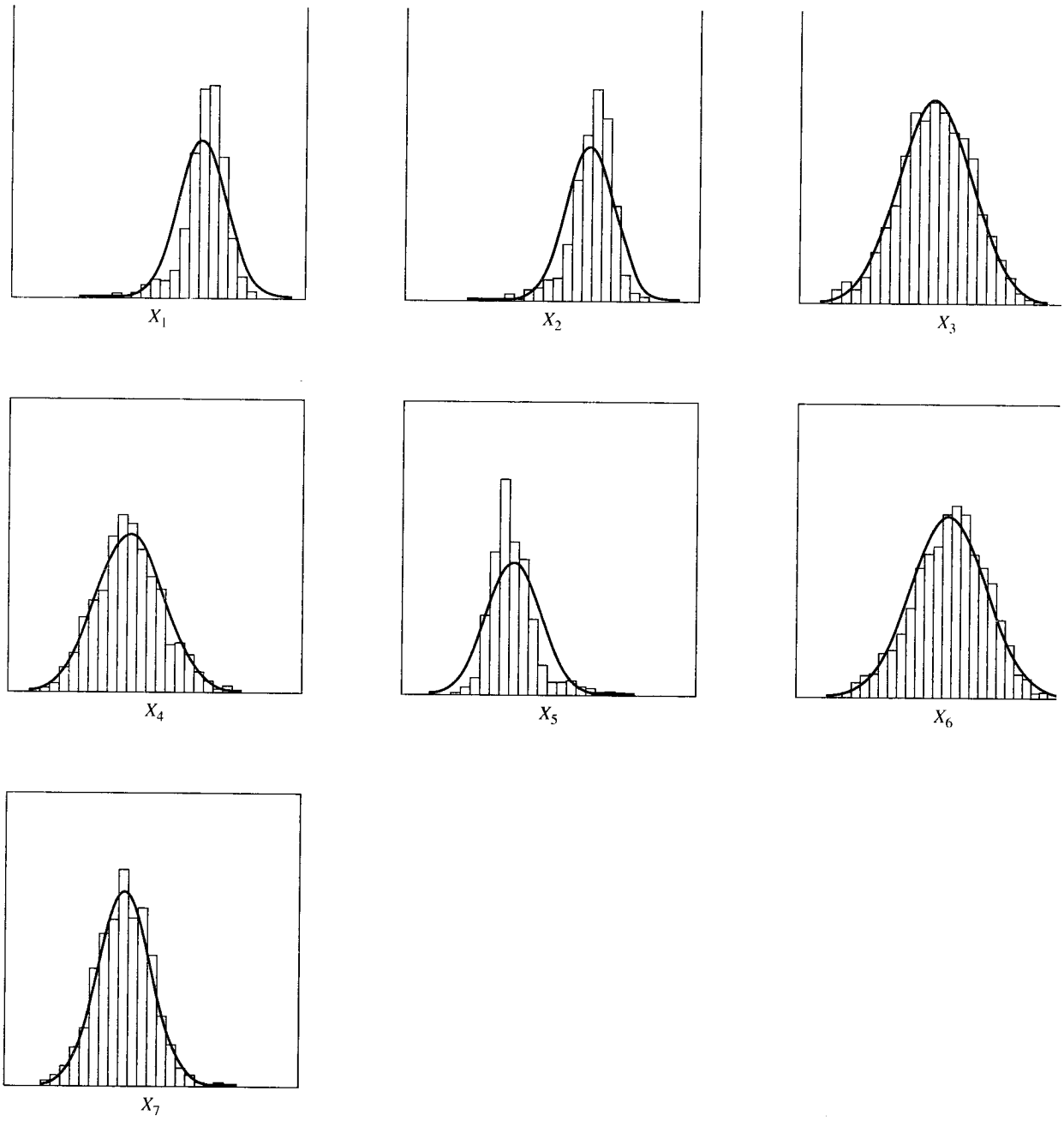


FIGURA 12.4. Distribuciones muestrales *bootstrap* para los siete coeficientes de regresión (de X_1 a X_7).

Resumen

Los procedimientos de remuestreo que hemos discutido ofrecen una perspectiva alternativa en una de las evaluaciones claves hecha en el análisis de los datos, la variabilidad del parámetro estimado. Ésta es la base de la hipótesis de contrastación y evaluación de significación estadística. Las técnicas de resmuestreo realzan la capacidad del investigador para examinar la distribución efectiva de los parámetros estimados en lugar de basarse en una distribución supuesta. Técnicas como éstas ofrecen un medio directo de «conocer los datos» y evitar todos los escollos habituales de ser muy dependientes de las técnicas estadísticas en lugar del juicio razonado del conocimiento de los datos.

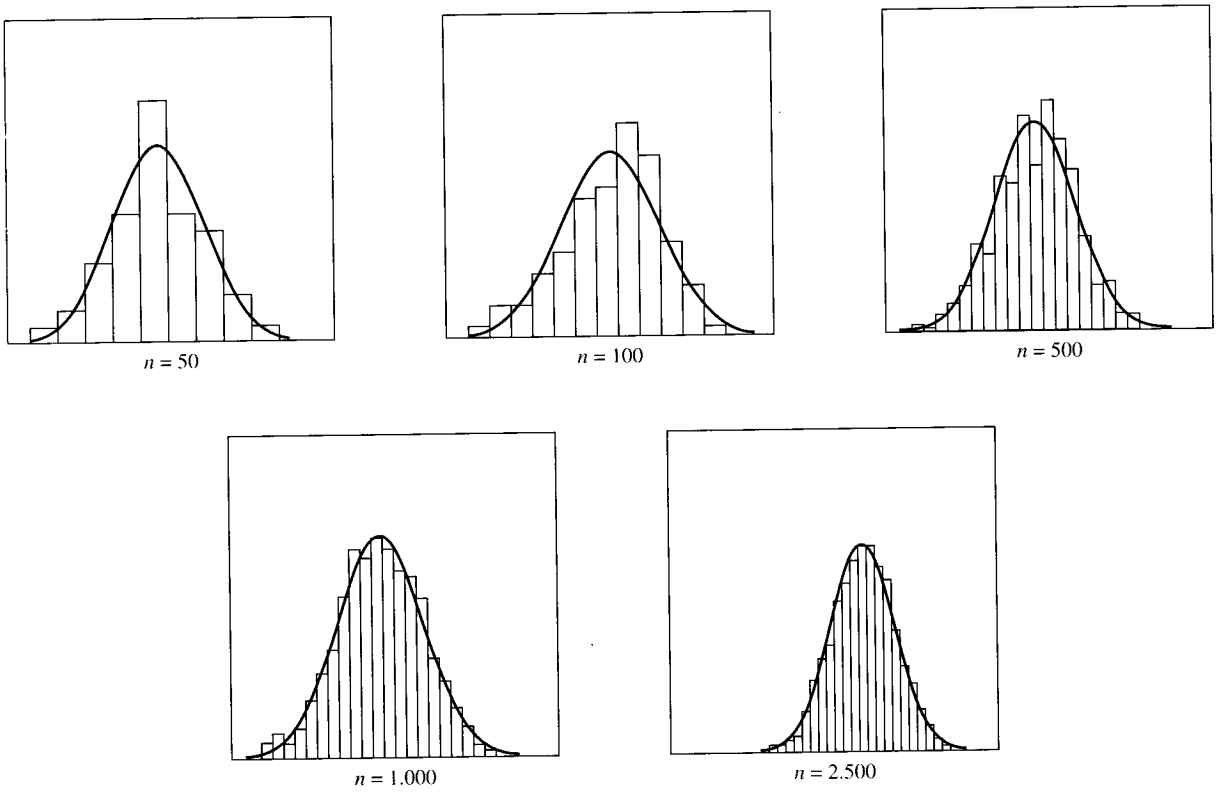


FIGURA 12.5. Cambios en la distribución muestral de *bootstrap* del coeficiente de regresión (X_3) para varios tamaños muestrales.

Resumen

Este capítulo final ha intentado ofrecer una perspectiva más amplia del análisis de los datos que la conseguida sólo por el estudio de las técnicas multivariantes. Nos encontramos en un período de intenso cambio. La información abunda y las organizaciones reconocen su creciente valor al tomar decisiones tácticas o estratégicas. Pero a medida que el ritmo de la toma de decisiones aumenta, los investigadores deben tener cada vez más en cuenta todas las técnicas disponibles y estar preparados para su aplicación. Una observación interesante es que muchas de estas técnicas y aspectos descritos en este capítulo evolucionan a partir de las necesidades identificadas primariamente en los campos aplicados del investigador multivariante. Esperamos que esto señale el comienzo de una era de mayor cooperación entre los usuarios orientados a la aplicación y los académicos para crear una verdadera colaboración en beneficio de ambas partes.

Cuestiones de revisión

1. Defina el papel que la extracción y almacenamiento de datos juega en la investigación actual.
2. Explique la lógica subyacente a los árboles de decisión y cómo pueden ser utilizados para tratar un problema de investigación.
3. Explique cómo se preparan y aprenden por sí mismas las redes neuronales. ¿Por qué no pueden hacer lo mismo las técnicas multivariantes?

4. ¿Cuáles son las principales diferencias entre las técnicas de reespecificación de *bootstrapping* y *jackknife*?
5. Describa cómo funciona un algoritmo genético. ¿Cuáles son sus principales ventajas y desventajas?
6. Describa las diferencias de los procedimientos mediante los cuales un método paramétrico y no paramétrico (reespecificación) estiman los intervalos de confianza.

Bibliografía

1. Balakrishanan, P. V., M. C. Cooper, V. S. Jacob, and P. A. Lewis (1994), «A Study of the Classification Capabilities of Neural Networks Using Unsupervised Learning: A Comparison with K-means Clustering.» *Psychometrika* 59: 509-25.
2. Banquin, R., and H. Edelstein, eds. (1996), *Planning and Designing the Data Warehouse.* Upper Saddle River, N.J.: Prentice Hall.
3. Berry, M., and G. Linoff (1997), *Data Mining Techniques for Marketing, Sales and Customer Support.* New York: Wiley.
4. Bigus, J. (1996), *Data Mining with Neural Networks.* New York: McGraw-Hill.
5. Bigus, J. (1996), *Data Mining with Neural Networks: Solving Business Problems-from Application Development to Decision Support.* New York: McGraw-Hill.
6. Boar, Bernard (1996), *Understanding Data Warehousing Strategically.* Lincroft, N.J.: NCR Corporation.
7. Chester, M. (1993), *Neural Networks: A Tutorial.* Upper Saddle River, N.J.: Prentice Hall.
8. Devlin, B. (1977), *Data Warehouse: From Architecture to Implementation.* New York: Addison-Wesley.
9. Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, vol. 38 of the CBSM-NSF Regional Conference Series in Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics.
10. Efron, B., and R. J. Tibshiran (1993), *An Introduction to the Bootstrap.* New York: Chapman and Hall.
11. Fausett, L. (1994), *Fundamental of Neural Networks: Architectures, Algorithms and Applications.* Upper Saddle River, N.J.: Prentice Hall.
12. Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (1996), *Advances in Knowledge Discovery and Data Mining.* Cambridge, Mass.: AAAI Press/MIT Press.
13. Groth, R. (1997), *Data Mining: A Hands-On Approach for Business Professionals.* Upper Saddle River, N.J.: Prentice Hall.
14. Hackathorn, D. (1995), «Reinventing Enterprise Systems via Data Warehousing.» The Data Warehousing Institute Annual Conference, Washington, D.C.
15. Hertz, J. (1991), *Introduction to the Theory of Neural Computing.* Reading, Mass.: Addison-Wesley.
16. Hinton, G. E. (1992), «How Neural Networks Learn from Experience.» *Scientific American* 267 (September): 144-51.
17. Holsheimer, M., and M. Kersten (1994), *Architectural Support for Data Mining.* Technical Report CS-R9429. Amsterdam: CWI.
18. Inmon, W. H. (1996), *Building the Data Warehouse.* London: QED Publishing Group.
19. Inmon, W. H., and R. D. Hackathorn (1994), *Using the Data Warehouse.* New York: Wiley.
20. Inmon, W. H., and J. D. Welch, and K. Glassey (1987), *Managing the Data Warehouse.* New York: Wiley.
21. Kimball, R. (1996), *The Data Warehouse Toolkit.* New York: Wiley.
22. Medsker, L., and J. Liebowitz (1994), *Design and Development of Expert Systems and Neural Networks.* New York: Macmillan.
23. Michalewicz, Z. (1994), *Genetic Algorithms* 1 *Data Structures* 5 *Evolution Programs.* New York: Springer-Verlag.
24. Michie, D., D. J. Spiegelhalter, and C. C. Taylor (1994), *Machine Learning, Neural and Statistical Classification.* New York: Ellis Horwood.
25. Mooney, C. Z., and R. D. Duval (1993), *Bootstrapping: A Nonparametric Approach to Statistical Inference.* Newbury Park, Calif.: Sage.
26. Piatetsky-Shapiro, G., and W. Frawley (1991), *Knowledge Discovery in Databases.* Cambridge, Mass.: AAAI Press/MIT Press.
27. Redman, T. C. (1996), *Data Quality for the Information Age.* New York: Artech House.
28. Simon, A. R. (1995), *Strategic Database Technology: Management for the Year 2000.* New York: Morgan Kaufman.
29. Smith, M. (1993), *Neural Networks for Statistical Modeling.* New York: Van Nostrand Reinhold.
30. Sprauge, R. H., and H. Watson (1994), *Decision Support for Management.* Upper Saddle River, N.J.: Prentice Hall.
31. Thomsen, Erik (1997), *OLAP Solutions: Building Multidimensional Information Systems.* New York: Wiley.
32. Tukey, J. W. (1958), «Bias and Confidence in Not Quite Large Samples.» *Annals of Mathematical Statistics* 29: 614.
33. Turban, E. (1995), *Decision Support Systems and Expert Systems.* Upper Saddle River, N.J.: Prentice Hall.
34. Ulman, J. D. (1989), *Principles of Database Systems.* Santa Clara, Calif.: Computer Science Press.
35. Weiss, K. M., and C. A. Kulikuwski (1991), *Computer Systems that Learn.* New York: Morgan Kaufman.

APÉNDICE A

Aplicaciones del análisis multivariante

Introducción

Tal y como hemos observado, para cada técnica multivariante se necesitan programas estadísticos que faciliten los cálculos y resuelvan todo tipo de problemas. También se menciona en muchas ocasiones a lo largo del libro, que la aceptación y uso generalizado del análisis multivariante se debe en gran medida a la disponibilidad de programas informáticos y al acceso a computadores tanto por parte de los investigadores académicos como de los analistas empresariales. Con el paso del tiempo, los investigadores han ido buscando programas nuevos y más sofisticados con el fin de mejorar y ampliar su «caja de herramientas» para el análisis multivariante. Afortunadamente, hoy en día se dispone incluso de las técnicas más sofisticadas mediante los paquetes estadísticos principales (por ejemplo, SPSS, SAS, SYSTAT o BMDP). Estos programas, que antes eran de dominio exclusivo de los computadores tipo mainframe, tienen hoy versiones basadas en un PC con las mismas capacidades que sus predecesores, los grandes computadores mainframe. Con el incremento continuado de potencia informática, el análisis que tan sólo hace unos años necesitaba un computador tipo mainframe se encuentra ahora disponible para cualquier usuario que simplemente tenga acceso a un computador personal relativamente nuevo.

En nuestro intento para adaptar el conocimiento operativo de las técnicas del análisis multivariante, este apéndice se dedica a proporcionar los comandos de control necesarios para ejecutar las técnicas ilustradas en los capítulos precedentes. Donde es posible, se proporcionan las instrucciones para los diferentes análisis, necesarias para los dos paquetes estadísticos más populares: SPSS y SAS. Estos paquetes estadísticos han alcanzado un uso generalizado en todo tipo de computador, desde los computadores tipo mainframe a los computadores personales. Afortunadamente, existen pocas diferencias, si es que existen, entre los comandos de control que se necesitan para cada tipo de computador.

Nuestros ejemplos estarán orientados a aplicaciones de computadores personales, pero sólo se necesitan pequeñas modificaciones para el uso en otros sistemas.

Es necesario señalar dos excepciones para el uso de SPSS o SAS en lo que se refiere a nuestro análisis. En primer lugar, diversos tipos de análisis multidimensional se llevan a cabo con un paquete muy accesible de MDS, que se encuentra disponible para el computador personal. Estos comandos de control se documentan en una sección aparte. La otra excepción es el área de SEM. Empleamos el popular programa LISREL, y los comandos de control necesarios para su ejecución se proporcionarán también en sección aparte. En ambos casos, se dispone del software necesario a precios asequibles.

Suponemos que el lector tiene algún conocimiento operativo de los programas utilizados. Una tutoría que explique las características generales de los paquetes estadísticos o los comandos necesarios para la operación general de los paquetes resulta fuera del alcance de este libro. Aunque estamos más familiarizados con el SPSS, se ha realizado un gran esfuerzo para proporcionar los comandos de control comparables en la medida de lo posible.

Es importante destacar que el uso de la sintaxis de comando se está reduciendo considerablemente con el comienzo de los programas de PC, especialmente los programas tipo Windows. La capacidad del usuario para especificar el análisis mediante una serie de menús facilita enormemente el uso de los programas a través de una amplia gama de menús. Ya no es necesario que el analista aprenda la sintaxis antes de llevar a cabo un análisis simple. En su lugar, puede que solamente para los casos más avanzados haya que manejarse la sintaxis de comando. Se proporciona la sintaxis de comando de estos programas con el fin de ilustrar los comandos reales para usuarios experimentados y para proporcionar directrices para los usuarios nuevos si surge la necesidad de utilizar la sintaxis de comando. En cada análisis, presentamos una serie de comandos de control anotados, donde se explican los comandos específicos y se detallan los comandos opcionales.

Los avances continuados en estos programas estadísticos requieren cierta actualización de los comandos de control conforme pasan los años. Incitamos a los investigadores interesados en visitar nuestra página Web en http://www.prenhall.com. Mantendremos los comandos de control actualizados además de cualquier sugerencia de otros investigadores implicados en estos u otros programas estadísticos.

Software utilizado

Joreskog, K. G., y D. Sorbom (1993), *LISREL VIII.* Mooresville, Ind.: Scientific Software, Inc.
SAS Institute, Inc. (1990), *SAS User's Guide: Basics, Version 6.* Cary, N.C.: SAS Institute, Inc.
SAS Institute, Inc. (1990), *SAS User's Guide: Statistics, Version 6.* Cary, N.C.: SAS Institute, Inc.
Smith, Scott M. (1989), *PC-MDS: A Multidimensional Statistics Package.* Provo, Utah: Brigham Young University.
SPSS, Inc. (1997), *SPSS for Windows, Version 7.5.* Chicago: SPSS, Inc.
SPSS, Inc. (1997), *SPSS for Windows, Missing Data Module.* Chicago: SPSS, Inc.
SPSS, Inc. (1993), *SPSS Categories Module.* Chicago: SPSS, Inc.
SPSS, Inc. (1993), SPSS Categories Module. Chicago: SPSS, Inc.

COMANDOS DE CONTROL SPSS para WINDOWS

CAPÍTULO 2: EXAMEN GRÁFICO DE LOS DATOS

Para crear el archivo de sistema SPSS

```
DATA LIST /ID 4-6 X1 10-12 X2 16-18
  X3 21-24 X4 28-30 X5 34-36 X6 40-42
  X7 45-48 X8 51 X9 54-57 X10 61-63
  X11 66 X12 69 X13 72 X14 75.
```

Identifica las variables y la localización de columnas.

```
FORMATS ID X1 X2 X4 X5 X6 X10 (F3.1).
FORMATS X3 X7 X9 (F4.1).
FORMATS X8 X11 X12 X13 X14 (F1.0).
```

Especifica el formato de cada variable.

```
VARIABLE LABELS ID 'ID'
  /X1 'Delivery Speed'
  /X2 'Price Level'
  /X3 'Price Flexibility'
  /X4 'Manufacturer Image'
  /X5 'Overall Service'
  /X6 'Salesforce Image'
  /X7 'Product Quality'
  /X8 'Firm Size'
  /X9 'Usage Level'
  /X10 'Satisfaction Level'
  /X11 'Specification Buying'
  /X12 'Structure of Procurement'
  /X13 'Industry Type'
  /X14 'Buying Situation'
```

Especifica una etiqueta para cada variable.

```
VALUE LABELS X8 0 'SMALL' 1 'LARGE'
  /X11 0 'SPECIFICATION BUYING'
       1 'TOTAL VALUE ANALYSIS'
/X12 0 'DECENTRALIZED' 1 'CENTRALIZED'
  /X13 0 'FIRM TYPE ONE' 1 'FIRM TYPE TWO'
  /X14 1 'NEW TASK' 2 'MODIFIED REBUY'
       3 'STRAIGHT REBUY'
```

Especifica una etiqueta para los valores de las variables.

```
BEGIN DATA
      1   4,1   0,6   6,9   4,7   2,4   2,3   5,2   0   32,0   4,2   1   0   1   1
      2   1,8   3,0   6,3   6,6   2,5   4,0   8,4   1   43,0   4,3   0   1   0   1
      3   3,4   5,2   5,7   6,0   4,3   2,7   8,2   1   48,0   5,2   0   1   1   2
      .   .     .     .     .     .     .     .     .   .      .     .   .   .   .
      .   .     .     .     .     .     .     .     .   .      .     .   .   .   .
```

Se facilita una lista completa de la base de datos al final de este apéndice.

```
      .   .     .     .     .     .     .     .     .   .      .     .   .   .   .
      .   .     .     .     .     .     .     .     .   .      .     .   .   .   .
     98   2,0   2,8   5,2   5,0   2,4   2,7   8,4   1   38,0   3,7   0   1   0   1
     99   3,1   2,2   6,7   6,8   2,6   2,9   8,4   1   42,0   4,3   0   1   0   1
    100   2,5   1,8   9,0   5,0   2,2   3,0   6,0   0   33,0   4,4   1   0   0   1
END DATA.
```

```
SAVE OUTFILE='HATCO.SAV'.
```

Guarda los datos como un archivo de sistema.

Análisis de datos ausentes

```
MVA
X1 X2 X3 X4 X5 X6 X7 X9 X10
  /ID = ID
```

Especifica las variables métricas para incluirlas en el análisis de datos ausentes.

```
/TTEST PROB PERCENT=5
```
Realiza el test-*t* de diferencias de media entre encuestados con datos válidos y ausentes.

```
/CROSSTAB PERCENT=5
```
Tabulación cruzada de cualquier variable categórica entre grupos formados por encuestados con datos válidos y ausentes.

```
/MISMATCH PERCENT=5
```
Matriz de porcentajes de casos que son mal emparejados en cuanto a parejas de variables (es decir, una variable tiene datos válidos/ausentes y otra variable tiene el contrario).

```
/DPATTERN DESCRIBE=X1 X2 X3 X4 X5 X6 X7 X9
 X10

/MPATTERN DESCRIBE=X1 X2 X3 X4 X5 X6 X7 X9
 X10
/TPATTERN PERCENT=1 DESCRIBE=X1 X2 X3 X4
 X5 X6 X7 X9 X10
```
Tres diferentes tipos de retratos de pautas de datos ausentes.

```
/LISTWISE
/PAIRWISE
```
Cálculo de la matriz de correlación entre matrices de variables utilizando supresión por pareja y por lista.

```
/EM (TOLERANCE=0.001 CONVERGENCE=0.0001
 ITERATIONS=200)
```
Emplea un procedimiento EM (expectativa-maximización) para imputar los valores y luego estimar las medias y las correlaciones.

```
/REGRESSION (TOLERANCE=0.001 FLIMIT=4.0
 ADDTYPE=RESIDUAL).
```
Emplea un procedimiento basado en la regresión para imputar los valores y luego estimar las medias y las correlaciones.

Estadísticos descriptivos

```
EXAMINE
 VARIABLES=X1 X2 X3 X4 X5 X6 X7 X9 X10

/PLOT NONE

/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
 /NOTOTAL.
```
Selecciona las variables para calcular los estadísticos descriptivos.

Cancela todos los gráficos.

Selecciona los estadísticos descriptivos y el uso de la supresión de pareja para tratar datos ausentes.

Contrastes de Homocedasticidad

```
EXAMINE
 VARIABLES=X1 X2 X3 X4 X5 X6 X7 X9 X10 BY
 X8 X11 X12 X13 X14 /ID=ID

/PLOT SPREADLEVEL(1)

/STATISTICS DESCRIPTIVES
 /CINTERVAL 95
 /MISSING LISTWISE
 /NOTOTAL.
```
Selecciona variables métricas (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_9, X_{10}) para compararlas con variables categóricas (X_8, X_{11}, X_{12}, X_{13}, X_{14}).

Realiza el análisis de «spread-level» y el contraste de Levene.

Selecciona estadísticos descriptivos y utiliza la supresión por listas para tratar datos ausentes.

CAPÍTULO 3: EL ANÁLISIS FACTORIAL

Análisis de componentes principales con la rotación de VARIMAX (sin X_5)

```
FACTOR
  /VARIABLES X1 X2 X3 X4 X6 X7

  /CRITERIA ITERATE(50)

  /FORMAT BLANK(0)

  /PRINT ALL
  /PLOT EIGEN ROTAT(1,2)

  /EXTRACTION PC

  /ROTATION VARIMAX.
```

Especifica el procedimiento de análisis factorial de X_1 hasta X_4, X_6 y X_7. Se omite X_5 después de un análisis inicial.

Especifica el número máximo de iteraciones para la solución del factor; por defecto es de 25.

Controla el formato de las matrices de los factores.
BLANC (0,30) no muestra ninguna ponderación de una variable menor a 0,30. En este ejemplo, se muestran todos los valores.

Imprime todos los estadísticos disponibles.

Muestra los autovalores en orden descendiente en un cuadro de caída. Genera un cuadro de la carga de los factores con los factores 1 y 2 en los ejes.

Especifica el método de extracción para los componentes principales.

Requiere el método de rotación disponible. Por defecto es VARIMAX. Otros métodos de rotación (EQUAMAX, QUARTIMAX,OBLIMIN) están disponibles con comandos adicionales/ ROTATION.

Análisis de componentes principales con Rotación oblicua (sin X_5)

La sintaxis para el análisis de componentes principales con la rotación oblicua permanece igual con la excepción de la rotación que se especifica como oblicua (OBLIMIN).

```
FACTOR
  /VARIABLES X1 X2 X3 X4 X6 X7
  /CRITERIA ITERATE(50)
  /FORMAT BLANK(0)
  /PRINT ALL
  /PLOT EIGEN ROTAT(1,2)
  /EXTRACTION PC
  /ROTATION OBLIMIN.
```

Validación del análisis de componentes principales mediante división de la muestra con la rotación VARIMAX

La muestra se divide en dos submuestras iguales de 50 encuestados. Entonces se estima de nuevo el modelo de factores.

```
SET SEED=34567.
```

Especifica la semilla de número aleatorio.

```
COMPUTE SPLIT=UNIFORM(1)>.52.
```

Calcula la variable SPLIT (de división) con una distribución uniforme entre 0 y 1.

```
EXECUTE.
SORT CASES BY SPLIT.
```

Reordena la secuencia de casos que se basan en los valores de la variable de división (SPLIT).

```
SPLIT FILE
```

Divide el archivo en dos subgrupos que se analizan por separado.

```
LAYERED BY SPLIT.
FACTOR
```

Misma sintaxis como el análisis de componentes principales con rotación VARIMAX arriba mencionada.

```
/VARIABLES X1 X2 X3 X4 X6 X7
  /CRITERIA ITERATE(50)
  /FORMAT BLANK(0)
/PRINT ALL
  /PLOT EIGEN ROTAT(1,2)
  /EXTRACTION PC
  /ROTATION VARIMAX.
```

Análisis de factores común (sin X_5)

La diferencia principal entre el análisis de componentes principales y el análisis factorial común es la especificación del método de extracción, que a partir de ahora es PAF. Para esta serie de datos, se omitió también la variable X_5.

```
FACTOR
  /VARIABLES X1 X2 X3 X4 X6 X7
  /PRINT INITIAL EXTRACTION ROTATION FSCORE
  /CRITERIA MINEIGEN(1) ITERATE(150)
  /EXTRACTION PAF

  /CRITERIA ITERATE(150)


  /ROTATION VARIMAX.
```

Variables para ser analizadas sin X_5.

Especifica el método de extracción factorial común (PAF).
Tiene que incrementar el número de iteraciones posibles en el análisis, para asegurar convergencia.
Requiere el método de rotación disponible.

CAPÍTULO 4: EL ANÁLISIS DE REGRESIÓN MÚLTIPLE

Regresión múltiple

```
REGRESSION

/DESCRIPTIVES ALL

/STATISTICS ALL

/CRITERIA=PIN(.05) POUT(.10)




/DEPENDENT X9

/METHOD = STEPWISE X1 X2 X3 X4 X5 X6 X7



/PARTIALPLOT ALL



/SCATTERPLOT=(*ZRESID,X9)

/RESIDUALS HIST(ZRESID) NORM(ZREZID) ID(ID)
```

Inicia el procedimiento de regresión.

Pide todos los estadísticos descriptivos.

Imprime todos los estadísticos de resumen.

Especifica los criterios estadísticos utilizados para formar la ecuación de regresión:
(1) PIN= probabilidad de F-para entrar.
(2) POUT= probabilidad de F-para salir.

Especifica la variable dependiente como X_9.

Especifica el método de selección de variable como un método por etapas y las variables para el análisis como X_1 a X_7.

Especifica los cuadros de regresión parcial utilizando todas las variables independientes.

Especifica las variables para los gráficos de dispersión.

Especifica el *output* de información sobre los casos atípicos, estadísticos,

```
                                        histogramas y cuadros de probabilidad
                                        normal. ID especifica el número id como la
                                        etiqueta para cuadros de casos o casos
                                        atípicos.

/CASEWISE ALL SRE MAH SDR COOK LEVER    Especifica la inclusión de todos los casos
                                        en el cuadro de casos y nombra las
                                        variables de diagnóstico para ser
                                        utilizadas. Se dispone de otras variables
                                        de diagnóstico.

/SAVE PRED ZPRED MAHAL COOK LEVER RESID Genera nuevas variables, que sirven como
  ZRESID SRESID DRESID SDRESID DFBETA    medidas de diagnosis para identificar
  SDBETA DFFIT SDFIT COVRATIO.          observaciones influyentes.
```

Validación por división de la muestra de la estimación por etapas

```
SET SPEED=34567.                        Especifica la semilla de número aleatorio.

COMPUTE SPLIT=UNIFORM(1)>.52.           Calcula la variable SPLIT (de división) con
                                        una distribución uniforme de entre 0 y 1.

EXECUTE.
SORT CASES BY SPLIT.                    Reordena la secuencia de casos que se basa
                                        en los valores de la variable SPLIT.

SPLIT FILE                              Divide el archivo en dos subgrupos, que se
                                        analizan por separado.

LAYERED BY SPLIT.
REGRESSION                              Inicia el procedimiento de regresión.
  /DESCRIPTIVES ALL                     Pide todos los estadísticos descriptivos.
  /STATISTICS ALL                       Imprime todos los estadísticos de resumen.
  /CRITERIA=PIN(.05) POUT(.10)          Especifica los criterios estadísticos que se
                                        utilizan para formar la ecuación de
                                        regresión:
                                        (1) PIN= probabilidad de F-para entrar.
                                        (2) POUT= probabilidad de F para salir.
  /DEPENDENT X9                         Especifica la variable dependiente como X9.
  /METHOD=STEPWISE X1 X2 X3 X4 X5 X6 X7. Especifica el método de selección de
                                        variables como un método por etapas y las
                                        variables para análisis como X1 a X7.
```

CAPÍTULO 5: EL ANÁLISIS DISCRIMINANTE MÚLTIPLE Y LA REGRESIÓN LOGÍSTICA

Análisis discriminante de dos grupos

```
SET SEED 54321.                         Especifica una semilla para generar
                                        números aleatorios que sirvan para crear
                                        una ampliación de la muestra.

COMPUTE RANDZ=UNIFORM(1) > .65.         Calcula la variable RANDZ con una
                                        distribución uniforme entre 0 y 1.

EXECUTE.
DISCRIMINANT                            Inicia el análisis discriminante y
  /GROUPS X11 (0 1)                     especifica la variable de agrupación como
                                        X11 con un rango de valores de 0 y 1.
  /VARIABLES X1 X2 X3 X4 X5 X6 X7       Especifica las variables predictor
                                        utilizadas.
  /SELECT=RANDZ(0)                      Selecciona los casos con RANDZ igual a 0
                                        para el uso en la estimación de modelo.
```

```
/ANALYSIS ALL
/METHOD MAHAL

/PIN=.05
/POUT=.10
/PRIORS SIZE



/HISTORY STEP END
/STATISTICS ALL
/PLOT=COMBINED SEPARATE MAP

/CLASSIFY=NONMISSING POOLED
/ROTATE=COEFF STRUCTURE.
```

Especifica el método para la selección de las variables para su inclusión
(a) PIN= probabilidad de F para entrar,
(b) POUT= probabilidad de F para borrar.
Especifica las probabilidades previas de calidad de miembro del grupo para ser igual a la proporción de muestra de los casos que realmente encajan en cada grupo.
Produce el informe del resumen final.
Imprime todas las estadísticas disponibles.
Gráficos de grupos combinados, gráficos de grupos separados y mapa territorial.
Especifica una matriz rotada y una matriz de estructura rotada.

Análisis discriminante de tres grupos

La única diferencia entre el análisis discriminante de dos grupos y tres grupos es la especificación de la variable de agrupación y su alcance.

```
DISCRIMINANT /GROUPS=X14(1,3)
```

Ahora X_{14} representa la variable de agrupación con valores de 1 a 3.

Análisis de regresión logística

```
SET SEED=123456.
```

Especifica una semilla para generar números aleatorios que sirvan para crear la ampliación de la muestra.

```
COMPUTE RANDZ=UNIFORM(1)>.60.
```

Computa la variable RANDZ con una distribución uniforme de entre 0 y 1.

```
LOGISTIC REGRESSION X11 WITH X1, X2, X3,
  X4, X5, X6, X7
```

Inicia una regresión logística con X_{11} como la variable dependiente y X_1 hasta X_7 como variables independientes.

```
/METHOD=FSTEP
```

Especifica una selección de variables por etapas.

```
/SELECT=RANDZ EQ 0
```

Selecciona casos con RANDZ igual a 0 para el uso en la estimación de modelo.

```
/PRINT=ALL
```

Imprime todo el *output* disponible.

```
/CRITERIA=ITERATE(50)
```

Especifica el número máximo de iteraciones en 50.

```
/CASEWISE=PRED PGROUP RESID SRESID
ZRESID LEVEL COOK DFBETA.
```

Especifica las variables de diagnóstico para ser utilizadas en las listas por casos.

CAPÍTULO 6: ANÁLISIS MULTIVARIANTE DE LA VARIANZA

Análisis multivariante de la varianza (2 grupos)

```
MANOVA
X9 X10 BY X11(0 1)
```

Especifica el procedimiento MANOVA con X_9 y X_{10} como las variables dependientes y X_{11} como la variable independiente con un rango de 0 a 1.

```
/PRINT CELLINFO (MEANS CORR COV)
  DESIGN (COLLINEARITY)
```

Especifica el output impreso:
(1) Información de las celdas de medias, de las matrices de correlación y matrices de varianza covarianza
(2) Diagnóstico de colinealidad para las matrices de diseño.

```
HOMOGENITY(BARTLETT COCHRAN BOXM)
```

(3) Contrastes de homogeneidad, Bartlett-Box, F y M de Box

```
SIGNIF(MULT UNIV STEPDOWN)
```

(4) Contrastes de significación de contrastes multivariantes para diferencias de grupo y contrastes de reducción.

```
SIGNIF(EFSIZE)
```

(5) Contraste de significación de efecto de tamaño para los contrastes univariantes y *t*.

```
PARAMETERS (ESTIM)
```

(6) Parámetros estimados, incluyendo los errores estándar, contrastes *t* e intervalos de confianza.

```
ERROR(CORR)
```

(7) Contrastes de significación para la igualdad de covarianzas.

```
/PLOT=ALL
```

Traza un gráfico normal y uno normal sin tendencias, y un boxplot para las variables dependientes.

```
/POWER T(.05) F(.05)
```

Pide que los valores de potencia observada sean calculados para un nivel de signficación de 0,05.

```
/METHOD=UNIQUE
```

Pide el método de separaración en grupos a partir de las sumas de los cuadradados que corresponden a una combinación de medias sin ponderación (única).

```
/ERROR WITHIN+RESIDUAL
/DESIGN.
```

Especifica la estructura del modelo y tiene que ser el último subcomando del modelo. Por defecto (tal y como se muestra) es el modelo factorial completo.

Análisis multivariante de la varianza (3 grupos)

La única diferencia entre un MANOVA de dos grupos y un MANOVA de tres grupos es la especificación de la variable independiente y su gama de valores, que ahora es X_{14} con un rango de 1 a 3.

```
MANOVA X9 X10 BY X14(1,3)
```

Especifica el procedimiento de MANOVA con X_9 y X_{10} como las variables dependientes y X_{14} como la variable independiente con un rango de 1 a 3.

Análisis multivariante de la varianza (2 factores)

La única diferencia entre un MANOVA de dos grupos y un MANOVA de dos factores es la especificación de las variables independientes y su rango de valores.

```
MANOVA X9 X10 BY X13 (0,1) X14(1,3)
```

Especifica el procedimiento MANOVA con X_9 y X_{10} como las variables dependientes y X_{13} y X_{14} como las variables independientes.

CAPÍTULO 7: ANÁLISIS DEL CONJUNTO

Diseño de los estímulos: Generación de un diseño factorial ortogonal y fraccionado

Tarjetas de control utilizadas para la generación automática de una serie

ortogonal de estímulos en el análisis de conjunto. Tiene que indicar la SEMILLA para su reproducción exacta.

```
ORTHOPLAN
  /FACTORS=
  MIXTURE 'Product Form' ('Premixed'
     'Concentrate' 'Powder')
  NUMAPP 'Num of Applic' ('50' '100' '200')
  GERMFREE 'Disinfectant' ('Yes' 'No')
  BIOPROT 'Biodegradable' ('No' 'Yes')
  PRICE 'Price/Applic' ('35 cents' '49
    cents' '79 cents')
  /HOLDOUT=4.

SAVE OUTFILE='CPLAN1.SAV'.
```

Utiliza el programa ORTHOPLAN para generar un diseño factorial ortogonal y fraccionado para cinco factores, tres factores de 3 niveles y dos factores de 2 niveles, etiquetando cada nivel para cada factor.

Cuatro estímulos adicionales para la ampliación de la muestra.

Guarda el plan generado para su uso posterior.

Diseño de los estímulos: Especificación de un diseño factorial ortogonal y fraccionado

Las tarjetas de control necesarias para reproducir el plan usado en el ejemplo HATCO. También ejemplo del método para incorporar el diseño de estímulos especificado en lugar de la generación de los mismos a partir del método mencionado arriba.

```
DATA LIST FREE
  /MIXTURE NUMAPP GERMFREE BIOPROT
  PRICE STATUS_CARD_.

BEGIN DATA.
    2,00      3,00      1,00      1,00      1,00      0         1
    3,00      3,00      1,00      1,00      1,00      0         2
    1,00      2,00      1,00      2,00      2,00      0         3
    3,00      3,00      1,00      2,00      2,00      0         4
    3,00      1,00      1,00      1,00      3,00      0         5
    2,00      3,00      2,00      2,00      3,00      0         6
    1,00      2,00      1,00      1,00      3,00      0         7
    1,00      3,00      1,00      1,00      2,00      0         8
    3,00      2,00      2,00      1,00      2,00      0         9
    2,00      1,00      1,00      1,00      2,00      0         10
    3,00      2,00      2,00      1,00      1,00      0         11
    2,00      2,00      1,00      1,00      3,00      0         12
    1,00      3,00      2,00      1,00      3,00      0         13
    1,00      1,00      1,00      1,00      1,00      0         14
    2,00      2,00      1,00      2,00      1,00      0         15
    1,00      1,00      2,00      2,00      1,00      0         16
    2,00      1,00      2,00      1,00      2,00      0         17
    3,00      1,00      1,00      2,00      3,00      0         18
    2,00      2,00      1,00      1,00      2,00      1         19
    3,00      2,00      2,00      2,00      1,00      1         20
    3,00      3,00      1,00      2,00      3,00      1         21
    2,00      1,00      2,00      2,00      1,00      1         22
    1,00      1,00      1,00      2,00      3,00      2         23
    2,00      3,00      1,00      2,00      2,00      2         24
    3,00      3,00      1,00      2,00      1,00      2         25
END DATA.
```

Define los factores y las especificaciones
STATUS: 0 uso para la estimación
1-ampliación de la muestra
2-datos del simulador de elección

Los niveles de cada factor que definen cada estímulo para el uso en el simulador de estimación, validación y de elección.

```
SAVE OUTFILE='CPLAN1.SAV'.
```

Guarda el plan generado para su uso posterior.

Impresión de tarjetas (Descripciones de perfil completo)

```
GET FILE='CPLAN1.SAV'.                          Plan ortogonal.

PLANCARDS
  /FACTOR=MIXTURE NUMAPP GERMFREE BIOPROT       Especifica los factores.
    PRICE
  /FORMAT BOTH                                  Genera tanto las tarjetas como las listas.
  /TITLE 'HYPOTHETICAL INDUSTRIAL
    CLEANSER )CARD'.                            Título que aparece en cada tarjeta de
                                                  estímulo. La TARJETA pone el número de
                                                  tarjeta en cada estímulo.
```

Estimación del modelo de análisis conjunto

Las tarjetas de control necesarias para (1) introducir los datos de preferencia proporcionados por los encuestados al evaluar los estímulos y (2) estimar el modelo de conjunto.

```
DATA LIST FREE/ QN PROD1 to PROD22.
BEGIN DATA.
  104  4  6  5  4  4  6  4  4  4  4  4  4  4  5  5  4  4  4  4  4  6  6
  107  6  3  5  2  3  1  1  6  6  6  7  4  1  6  6  6  6  1  7  7  1  7
       The complete data set is listed at the end of the appendix
  417  5  5  2  5  1  1  2  2  2  2  1  1  1  4  6  3  2  1  4  4  2  3
  418  6  7  2  7  1  4  1  3  6  4  7  1  5  2  1  2  4  2  3  7  5  7
END DATA.

CONJOINT PLAN='CPLAN1.SAV'                      Recupera el plan ortogonal.

  /FACTORS=                                     Selecciona los factores en la estimación de
  MIXTURE 'Product Form' ('Premixed'              conjunto.
    'Concentrate' 'Powder')
  NUMAPP 'Number of Applications' ('50'
    '100' '200')
  GERMFREE 'Disinfectant' ('Yes' 'No')
  BIOPROT 'Biodegradable' ('No' 'Yes')
  PRICE 'Price per Application' ('35 cents'
    '49 cents' '79 cents')
  /SUBJECT=QN                                   Selecciona la variable QN como sujeto ID.
  /SCORE=PROD1 PROD2 PROD3 PROD4 PROD5          Especifica la respuesta de preferencia de
    PROD6 PROD7 PROD8 PROD9 PROD10 PROD11       las variables. Tienen que ser puestas en la
    PROD12 PROD13 PROD14 PROD15 PROD16          lista por orden de perfil en el diseño
    PROD17 PROD18 PROD19 PROD20 PROD21          ortogonal.
    PROD22
  /UTILITY='UTIL.SAV'.                          Guarda las estimaciones de las funciones
                                                  parciales.
```

CAPÍTULO 8: CORRELACIÓN CANÓNICA

SPSS no tiene un procedimiento separado para el análisis canónico pero se puede llevar a cabo mediante el procedimiento MANOVA con estos comandos.

```
MANOVA X9 X10 WITH X1 TO X7
  /PRINT=ERROR (SSCP COV COR) SIGNIF             Imprime matrices de error y contrastes de
    (HYPOTH EIGEN DIMENR)                          significación
  /DISCRIM=RAW STAN ESTIM COR ALPHA(1.0)         Genera los coeficientes básicos y
                                                   estandarizados de la función discriminante
                                                   (RAW y STAN), las estimaciones de efecto
                                                   (ESTIM) y las correlaciones entre las
                                                   variables dependientes y canónicas (COR).
```

```
                                        Se anotan todas las funciones discriminantes
                                        (ALPHA (1,0)) para el análisis
                                        discriminante canónico
  /RESIDUALS=CASEWISE PLOT              Resultados y gráficos de valores por casos y
  /DESIGN.                              residuos.
```

CAPÍTULO 9: EL ANÁLISIS CLUSTER

Jerárquico

```
PROXIMITIES                             Computa la proximidad de observaciones entre
  X1 X2 X3 X4 X5 X6 X7                  sí a través de las variables (X1 hasta X7)
  /MATRIX OUT ('C:\WINDOWS\TEMP\SPSSCLUS.TMP')  introducidas en el procedimiento cluster.
  /VIEW=CASE
  /MEASURE=SEUCLID                      Especifica la medida de distancia utilizada,
                                        distancia Euclídea (por defecto).
  /PRINT=NONE
  /STANDARDIZE=NONE.
CLUSTER
  /MATRIX IN ('C:\WINDOWS\TEMP\SPSSCLUS.TMP')   Recupera el archivo de proximidades.
  /METHOD=WARD(WCLUS)                   Especifica el método cluster: en este caso el
                                        Método de Ward, que requiere distancias
                                        Euclídeas al cuadrado. WCLUS especifica un
                                        nombre raíz para guardar los clusters
                                        (véase el próximo comando).
  /PRINT=SCHEDUAL DISTANCE CLUSTER(2,5) Imprime los coeficientes de aglomeración y
                                        la matriz de distancia.
                                        Imprime la pertenencia al cluster para cada
                                        caso, de 2 a 5 clusters.
  /PLOT=DENDROGRAM VICICLE              Gráfico basado en el dendrograma.
  /SAVE=CLUSTER(2,5).                   Guarda cada pertenencia de cluster de cada
                                        caso para la solución cluster de dos,
                                        tres, cuatro y cinco. Las nuevas variables
                                        son WCLUS5, WCLUS4, WCLUS3 y WCLUS2.

ERASE FILE=                             Elimina el archivo de proximidades.
  'C:\WINDOWS\TEMP\SPSSCLUS.TMP'.
```

No jerárquico (Centroides de cada cluster especificados a priori)

```
QUICK CLUSTER X1 X2 X3 X4 X5 X6 X7      Agrupación no jerárquico de X1 a X7.
  /INITIAL=(4.46 1.576 8.9 4.926 2.992 2.51   Proporciona los centroides iniciales,
    5.904 2.57 3.152 6.888 5.57 2.84 2.82     leyendo los centroides de cluster para el
    8.038)                                    grupo 1, después grupo 2 sobre las
                                              variables X1, X2..., X7. En este ejemplo,
                                              el centroide para el grupo 1 sobre X1 es
                                              4,46 mientras que el grupo 2 tiene un
                                              valor medio sobre X1 de 2,57.
  /CRITERIA=CLUSTERS (2)                Especifica la formación de dos cluster.
  /PRINT=CLUSTER ANOVA                  Imprime el contraste ANOVA para las
                                        diferencias de cada variable a través de
                                        los cluster.
  /SAVE=CLUSTER(NHCLUS).                Guarda la pertenencia al cluster en
                                        NHCLUS.
```

No jerárquico (Selección aleatoria de puntos de semilla de cluster)

```
SET SEED 345678.                        Especifica los centroides de forma
                                        aleatoria.

QUICK CLUSTER X1 X2 X3 X4 X5 X6 X7      Agrupación no jerárquica de X1 a X7.
```

```
/CRITERIA=CLUSTERS (2) NOINITIAL
```
Especifica la formación de dos clusters de centros de cluster iniciales seleccionados aleatoriamente.

```
/PRINT=CLUSTER ANOVA INITIAL
```
Imprime el contraste ANOVA para las diferencias de cada variable a través de los cluster.

```
/SAVE=CLUSTER(NHRCLUS).
```
Igual que arriba, la variable es ahora NHRCLUS.

CAPÍTULO 10: ESCALA MULTIDIMENSIONAL

Escala multidimensional (INDSCAL)

```
ALSCAL
```
Selecciona el procedimiento ALSCAL para el procedimiento de análisis multidimensional clásico.

```
VARIABLES=var1 var2 var3 var4 var5 var6
  var7 var8 var9 var10
```
Define 10 clasificaciones de similitud.

```
/SHAPE=symmetric
```
Especifica que se usa una matriz simétrica por lo que no sea necesario poner las distancias en la porción superior de la matriz.

```
/INPUT ROWS(10)
```
Define el número de filas.

```
/LEVEL=ratio (1)
```
Especifica el nivel métrico de análisis debido a los datos de clasificación utilizados como *input*.

```
/CONDITION=MATRIX
```
Especifica que cada uno de los datos de los encuestados es único.

```
/MODEL=INDSCAL
```
Selecciona el procedimiento INDSCAL (Modelo de escala de diferencias individuales).

```
/CRITERIA=CONVERGE(.001) STRESSMIN(.001)
  ITER(50) CUTOFF(0) DIMENS(1,5)
```
Parámetros de estimación por defecto.

```
/PLOT=DEFAULT ALL
```
Selecciona todos los gráficos posibles.

```
/PRINT=HEADER.
```

Ajuste de propiedad (PROFIT) y protección de referencia (PREFMAP)

Veánse los comandos de control anotados en la sección PC-MDS

Análisis de correspondencias

```
DATA LIST FREE/ HATCO FIRM_A FIRM_B FIRM_C
  FIRM_D FIRM_E FIRM_F FIRM_G FIRM_H FIRM_I.
```
Define los atributos de clasificación de las empresas.

```
BEGIN DATA.
  4   3   1  13   9   6   3  18   2  10
 15  16  15  11  11  14  16  12  14  14
 15  14   6   4   4  15  14  13   7  13
 16  13   8  13   9  17  15  16   6  12
 14  14  10  11  11  14  12  13  10  14
  7  18  13   4   9  16  14   5   4  16
  6   6  14  10  11   8   7   4  14   4
 15  18   9   2   3  15  16   7   8   8
END DATA.
```
Las entradas individuales en la matriz son el número de veces que se clasifica una empresa con relación a su posesión de un atributo específico.

```
ANACOR

  TABLE=ALL(8,10)
  /DIMENSION=2

  /NORMALIZATION CANONICAL

  /PRINT TABLE SCORES CONTRIBUTIONS PROFILES
    (PERMUTATION
  /VARIANCES ROWS COLUMNS SINGULAR
  /PLOT ROWS COLUMNS JOINT NDIM(ALL,MAX).
```

Realiza un análisis de correspondencias basado en la matriz de tabulación cruzada.
Especifica el número de filas y columnas.
Especifica el número de dimensiones a calcular
Especifica el método de normalización de las puntuaciones de filas y columnas.
Imprime los estadísticos de correspondencia seleccionados.

Gráfico de las puntuaciones de filas y columnas separadas (FILAS COLUMNAS) y combinadas (JOINT) además de las dos dimensiones (NDIM).

COMANDOS DE CONTROL SAS

CAPÍTULO 2: EXÁMEN GRÁFICO DE LOS DATOS

Creando el archivo de datos SAS

```
DATA HATCO;
```

Especifica un nombre temporal para el archivo de datos.

```
INPUT ID 4-6 X1 10-12 X2 16-18 X3 21-24
  X4 28-30 X5 34-36 X6 40-42 X7 45-48
  X8 51 X9 54-57 X10 61-63
  X11 66 X12 69 X13 72 X14 75;
```

Identifica variables y localización de columnas.

```
LABEL ID 'ID'
X1 'Delivery Speed'
X2 'Price Level'
X3 'Price Flexibility'
X4 'Manufacturer Image'
X5 'Overall Service'
X6 'Salesforce Image'
X7 'Product Quality'
X8 'Firm Size'
X9 'Usage Level'
X10 'Satisfaction Level'
X11 'Specification Buying'
X12 'Structure of Procurement'
X13 'Industry Type'
X14 'Buying Situation';
```

Especifica una variable para cada etiqueta.

```
CARDS:
   1   4,1   0,6   6,9   4,7   2,4   2,3   5,2   0   32,0   4,2   1   0   1   1
   2   1,8   3,0   6,3   6,6   2,5   4,0   8,4   1   43,0   4,3   0   1   0   1
   .    .     .     .     .     .     .     .    .     .      .    .   .   .   .
   .    .     .     .     .     .     .     .    .     .      .    .   .   .   .
  A complete listing of the dataset is provided at the end of this Appendix
   .    .     .     .     .     .     .     .    .     .      .    .   .   .   .
   .    .     .     .     .     .     .     .    .     .      .    .   .   .   .
  99   3,1   2,2   6,7   6,8   2,6   2,9   8,4   1   42,0   4,3   0   1   0   1
 100   2,5   1,8   9,0   5,0   2,2   3,0   6,0   0   33,0   4,4   1   0   0   1
RUN;
```

Estsadísticos descriptivos

```
PROC UNIVARIATE DATA=HATCO NORMAL
PLOT;
VAR X1 X2 X3 X4 X5 X6 X7 X9 X10;
RUN;
```

Realiza los contrastes necesarios para probar la normalidad de las variables e identifica el número de valores ausentes por cada variable.

CAPÍTULO 3: ANÁLISIS FACTORIAL

Análisis de componentes principales

```
PROC FACTOR CORR MSA SCREE;
ROTATE=VARIMAX;
VAR X1-X4 X6 X7;
RUN;
```

Inicia el procedimiento de factores con una rotación VARIMAX que genera las correlaciones, MSAs y un cuadro de caída.

```
PROC FACTOR CORR MSA SCREE
ROTATE=PROMAX
VAR X1-X4 X6 X7;
RUN;
```

Inicia el procedimiento del análisis factorial con una rotación PROMAX con el resultado de correlaciones, MSAs y un cuadro de caída.

Análisis factorial común

```
PROC FACTOR METHOD=PRINIT CORR MSA
SCREE ROTATE=VARIMAX;
X1-X4 X6 X7;
RUN;
```

Igual que el análisis de componentes principales con la excepción de que el VAR método se especifica como un análisis factorial común.

CAPÍTULO 4: ANÁLISIS DE REGRESIÓN MÚLTIPLE

Regresión múltiple

```
PROC REG;
```

Inicia el procedimiento de regresión.

```
MODEL X9=X1-X7
  /ALL SELECTION=STEPWISE PARTIAL;
```

Identifica el modelo para ser usado: X_9 como la variable dependiente, X_1 a X_7 como variables independientes. Todas las estadísticas que se dan con el procedimiento de entrada por etapas. Requiere cuadros de apalancamiento de regresión parcial para cada variable independiente.

```
PLOT R.*P.;
```

Especifica las variables para representar en gráficos residuos y valores previstos.

```
RUN;
```

CAPÍTULO 5: ANÁLISIS DISCRIMINANTE MÚLTIPLE

Análisis discriminante de 2 grupos

```
PROC DISCRIM METHOD=NORMAL;
POOL=YES LIST CROSSVALIDATE;
```

Inicia el procedimiento de análisis discriminante con todas las estadísticas y validación.

```
CLASS X11;
```

Especifica X_{11} como una variable dependiente.

```
VAR X1-X7;
```

Especifica las variables predictor de X_1 a X_7.

```
RUN;
```

Análisis discriminante de 3 grupos

La única modificación necesaria para un análisis discriminante de 3 grupos es la identificación de una nueva variable de clasificación, X_{14}, una variable de tres grupos.

```
CLASS X14;
```

Selecciona X_{14} como una variable de clasificación.

Análisis de regresión logística

```
PROC CATMOD;
```

Inicia la modelación de datos categóricos.

```
DIRECT X1-X7;
```

Especifica las variables independientes que contienen valores de la matriz.

```
MODEL X11=X1-X7/ML CORRB FREQ ONEWAY
PRED=PROB NOGLS XPX;
```

Especifica la variable dependiente como X_{11}, y las variables independientes, X_1 a X_7, además de *output* opcional adicional. Para usarlo como regresión logística, es necesario solicitar ML y NOGLS, que utiliza estimaciones de máximo-verosimilitud y suprime el cálculo de mínimos cuadrados generalizados (ponderados).

```
RUN;
```

CAPÍTULO 6: ANÁLISIS MULTIVARIANTE DE LA VARIANZA

Análisis multivariante de la varianza (de 2 grupos)

```
PROC GLM;
CLASS X11;
MODEL X9 X10 = X11;
MEANS X11 / BON SNK TUKEY;
MANOVA H=X11 / SUMMARY;
RUN;
```

Inicia el modelo lineal general (GLM).

Selecciona X_{11} como la variable de clasificación.

Identifica el modelo MANOVA con X_{11} como la variable independiente y X_9, X_{10} como las variables dependientes.

Solicita la media para cada nivel de X_{11} con los contrastes *t* de Bonferroni, los contrastes de gama múltiple y el contraste de rango studentizado de Tukey sobre los efectos principales.

Especifica el efecto empleado igual que las matrices de hipótesis y las tablas de ANOVA para cada variable dependiente.

Análisis multivariante de la varianza (de 3 grupos)

La única modificación necesaria para un análisis de MANOVA de tres grupos es la variable de clasificación, X_{14}, una variable de 3 grupos.

```
CLASS X14;
```

Selecciona X_{14} como una variable de clasificación.

Análisis multivariante de la varianza (de 2 factores)

La única modificación necesaria para un análisis MANOVA de 2 factores son las variables de clasificación, X_{13} y X_{14}.

```
CLASS X13 X14;
```

Identifica las variables de clasificación como X_{13} y X_{14}.

CAPÍTULO 8: CORRELACIÓN CANÓNICA

```
PROC CANCORR ALL;
VAR X9 X10;
WITH X1-X7;
RUN;
```

Inicia el procedimiento de correlación canónica con todo el *output* adicional.

Selecciona X_9 y X_{10} como variables dependientes.

Selecciona X_1 a X_7 como las variables predictor.

CAPÍTULO 9: ANÁLISIS CLUSTER

Jerárquico

```
PROC CLUSTER M=WARD PSEUDO OUT SEED2;
VAR X1-X7;
```

Inicia el procedimiento rápido de cluster, especifica un número máximo de clusters (en este caso dos).

Identifica las variables para el análisis cluster como X_1 a X_7.

```
PROC TREE N=5;
```

Inicia el procedimiento de árbol con el máximo número de clusters para ser incluidos en el diagrama como 5. Al igual que un dendrograma.

```
PROC PRINT DATA=SEED2;
PROC FASTCLUS MAXC=2 MEAN=NEW OUT=TWOCLUS;
RUN;
```

Especifica el archivo con centros cluster, identifica el archivo nuevo temporal para ser guardado como TWOCLUS.

No jerárquico (Centroides especificados a priori)

```
PROC FASTCLUS SEED=NEW MAXC=2 OUT=TWOCLUS
```

Inicia el procedimiento cluster utilizando la media del cluster previamente generado como los puntos semilla.

```
VAR X1-X7;
RUN;
```

Identifica las variables para ser X_1 a X_7.

```
PROC ANOVA
CLASS CLUSTER
MODEL X1-X7=CLUSTER;

RUN;
```

Utiliza el procedimiento ANOVA con los dos clusters y las variables independientes de X_1 a X_7. Comprobando así las diferencias significativas entre los clusters sobre las variables utilizadas.

No jerárquico (Centroides seleccionados aleatoriamente)

```
PROC FASTCLUS MAXC=2 OUT=TWOCLUS
  REPLACE=RANDOM
```

Igual que arriba pero con una selección aleatoria de los centros de cluster iniciales.

```
VAR X1-X7;
```

```
PROC ANOVA;
CLASS CLUSTER;
MODEL X1-X7=CLUSTER;
RUN;
```

Igual que arriba.

COMANDOS DE CONTROL LISREL VIII

CAPÍTULO 11: MODELOS DE ECUACIONES ESTRUCTURALES

Análisis factorial confirmatorio: Especificación del modelo inicial

```
CONFIRMATORY FACTOR ANALYSIS
```
Título de la tarjeta.

```
DA NI=7 NO=100 MA=KM
```
Especifica el archivo de datos para el número de variables (7), tamaño muestral (100) y tipo de datos (correlación =KM).

```
KM FU FI=C:\HATCO.COR FO=5
(7F9.0)
```
Lee archivo de datos del disco.

```
SELECT
  1 2 3 4 6 7 /
```
Selecciona correlaciones para el análisis de la matriz entera. Es necesario anotar que se omite la variable 5.

```
MO NX=6 NK=2 PH=ST TD=SY,FI
```
La tarjeta del modelo define el número de los indicadores exógenos (6), número de constructos exógenos (2), y las características de las matrices asociadas.

```
LA
'DelvSpd' 'PriceLvl' 'PriceFlx'
'MfgImage' 'Service' 'SalesImg' 'Quality'
```
Etiquetas para variables en la matriz de *input*.

```
LK
'Strategy' 'Image'
```
Etiquetas para los constructos exógenos.

```
PA LX
1 (1 0)
1 (1 0)
1 (1 0)
1 (0 1)
1 (0 1)
1 (1 0)
```
Matriz de pauta especificando las cargas de los indicadores sobre los constructos exógenos. Se sugiere este formato porque corresponde directamente con el formato familiar del análisis factorial y las discusiones de texto.

```
FR TD(1,1) TD (2,2) TD(3,3) TD(4,4) TD(5,5)
  TD(6,6)
```
"Libera" los términos de error de indicador para la estimación.

```
OU SS TV RS MI
```
Tarjeta de *output*: solicita una solución estandarizada, valores-*t*, residuos e índices de modificación.

Análisis factorial confirmatorio: Reespecificación de modelo

```
CONFIRMATORY FACTOR ANALYSIS
DA NI=7 NO=100 MA=KM
KM FU FI=C:\HATCO.COR FO=5
(7F9.0)
SELECT
  1 2 3 4 6 7 /
MO NX=6 NK=2 PH=ST TD=SY,FI

LA
'DelvSpd' 'PriceLvl' 'PriceFlx' 'MfgImage'
'Service' 'SalesImg' 'Quality'
LK
'Strategy' 'Image'
PA LX
```
Tarjetas igual que en el modelo previo.

```
1 (1 0)
1 (1 0)
1 (1 0)
1 (0 1)
1 (0 1)
1 (1 0)
FR TD(1,1) TD (2,2) TD(3,3) TD(5,5) TD(6,6)

VA .005 TD(4,4)                                  Especifica la varianza de error de la
                                                   variable 4 para ser 0,005 como solución
                                                   del caso Heywood.

OU SS TV RS MI                                   Igual que en el modelo anterior.
```

Análisis de factor confirmatorio: Estimación del modelo nulo

```
CONFIRMATORY FACTOR ANALYSIS - NULL MODEL
DA NI=7 NO=100 MA=KM
KM FU FI=C:\HATCO.COR FO=5
(7F9.0)
SELECT                                           Tarjetas igual que con modelos previos.
  1 2 3 4 6 7 /

MO NX=6 NK=1 PH=ST TD=SY,FI                      Especifica el constructo exógeno simple para
                                                   el modelo nulo.
LA

'DelvSpd' 'PriceLvl' 'PriceFlx' 'MgfImage'       Igual que anteriormente.
'Service' 'SalesImg' 'Quality'

LK                                               Etiqueta el factor simple como el modelo
'Null Mod'                                         nulo.


PA LX                                            Especifica no carga para los indicadores
                                                   (véase abajo).
6 (0)
VA 1.0 LX(1,1) LX(2,1) LX(3,1) LX(4,1)           Especifica que las cargas del constructo
VA 1.0 LX(5,1) LX(6,1)                             para todos los indicadores igualen a 1,0
FR TD(1,1) TD (2,2) TD(3,3) TD(4,4)                (sin medición de error) y libera los
FR TD(5,5) TD(6,6)                                 términos de error para la estimación.

OU SE TV RS SS MI
```

Modelo de ecuación estructural (Modelo de trayectoria): Estimación del modelo

```
CAUSAL MODEL WITH MULTIPLE INDICATORS
DA NI=15 NO=136 MA=KM                            Especifica un archivo de correlación con 15
KM FU FILE=C:\STRUC1.COR FO=5                      variables y tamaño muestral de 136 para
(8F6.4/7F6.4)                                      ser leídos del disco.

MO NX=13 NK=3 NY=2 NE=2 GA=FU,FI                 El modelo consta de 13 indicadores para tres
  PS=SY,FI C BE=FU,FI TE=SY,FI PH=SY,FR            constructos exógenos y dos constructos
                                                   endógenos con un indicador cada uno
                                                   (total=15). También se definen las matrices
                                                   asociadas.

LA
'USAGE' 'SATISFAC' 'PRODQUAL' 'INVACCUR'
'TECHSUPT' 'NEWPROD' 'DELIVERY'
'MKTLEADR' 'PRDVALUE' 'LOWPRICE' 'NEGOTIAT'      Etiquetas para las variables en la matriz
'MUTUALTY' 'INTEGRTY' 'FLEXBLTY' 'PROBRES'         de input.
```

```
LK                                                    Etiquetas del constructo exógeno.
'FIRMPROD' 'PRICEFAC' 'RELATFAC'

LE                                                    Etiquetas del constructo endógeno.
'USAGE' 'SATISFAC'

PA LX
1 (0 0 0)
1 (1 0 0)
1 (1 0 0)
1 (1 0 0)                                             Especificación del modelo de medición para
1 (1 0 0)                                               V1, indicadores exógenos. Anótese que V1,
1 (1 0 0)                                               V7 y V10 no tienen cargas porque cada uno
1 (0 0 0)                                               será ajustado a 1,0 para controlar la no
1 (0 1 0)                                               varianza de escala (véase abajo).
1 (0 1 0)
1 (0 0 0)
1 (0 0 1)
1 (0 0 1)
1 (0 0 1)
PA GA
1 (1 -1 1)                                            Especifica las coeficientes exógenas para
1 (0 0 0)                                               las ecuaciones estructurales.
PA BE
1 (0 0)                                               Especifica las coeficientes endógenas para
1 (1 0)                                                 las ecuaciones estructurales.
PA PHI
1                                                     Correlaciones entre los constructos exógenos.
1 1
1 1 1
PA PS
1                                                     Ninguna correlación entre los constructos
0 1                                                     endógenos.

VA 1 LX(1,1) LX(7,2) LX(10,3) LY(1,1)                 Ajusta las cargas del indicador a 1,0 para
  LY(2,2)                                               controlar la no varianza de escala.

VA 0.00 TE(2,2) TE(1,1)                               Ajusta el error de medición a 0 para
                                                        indicadores de artículos simples para los
                                                        constructos endógenos.

OU SE TV RS SS MI AD=OFF                              Especifica el output.
```

Modelos de ecuaciones estructurales (modelo de trayectoria): Estimación del modelo nulo

```
CAUSAL MODEL WITH MULTIPLE INDICATORS—NULL MODEL
DA NI=15 NO=136 MA=KM
KM FU FILE=C:\STRUC1.COR FO=5
(8F6.4/7F6.4)
SELECT
  1 2 3 4 5 6 7 8 9 10 11 12 13 14 15/              Igual que anteriormente con la excepción de
MO NX=15 NK=1 TD=DI,FR PH=SY,FR                       las nuevas variables.
LA
'USAGE' 'SATISFAC' 'PRODQUAL' 'INVACCUR' 'TECHSUPT'
'NEWPROD' 'DELIVERY' 'MKTLEADR' 'PRDVALUE' 'LOWPRICE'
'NEGOTIAT' 'MUTUALTY' 'INTEGRTY' 'FLEXBLTY' 'PROBRES'
PA LX
15 (1)
OU SE TV RS SS MI
```

Modelos de ecuaciones estructurales (modelo de trayectoria): Modelo de competición (COMPMOD1)

```
CAUSAL MODEL WITH MULTIPLE INDICATORS — COMPMOD1
DA NI=15 NO=136 MA=KM
```

```
KM FU FILE=C:\STRUC1.COR FO=5
(8F6.4/7F6.4)
MO NX=13 NK=3 NY=2 NE=2 GA=FU,FI C
  PS=SY,FI BE=FU,FI TE=SY,FI PH=SY,FR
LA
'USAGE' 'SATISFAC' 'PRODQUAL' 'INVACCUR' 'TECHSUPT'
'NEWPROD' 'DELIVERY' 'MKTLEADR' 'PRDVALUE' 'LOWPRICE'
'NEGOTIAT' 'MUTUALTY' 'INTEGRTY' 'FLEXBLTY' 'PROBRES'
LK
'FIRMPROD' 'PRICEFAC' 'RELATFAC'
LE
'USAGE' 'SATISFAC'
PA LX
1 (0 0 0)
1 (1 0 0)
1 (1 0 0)                                    Igual que modelo estructural anterior.
1 (1 0 0)
1 (1 0 0)
1 (1 0 0)
1 (0 0 0)
1 (0 1 0)
1 (0 1 0)
1 (0 0 0)
1 (0 0 1)
1 (0 0 1)
1 (0 0 1)
PA GA
1 (1 1 1)                                    Especifica que los constructos exógenos son
1 (1 1 1)                                      construcciones que ahora se relacionan con
PA BE                                          todas las construcciones endógenas en las
1 (0 0)                                        ecuaciones estructurales.
1 (1 0)
PA PHI                                       Igual que el modelo anterior.
1
1 1
1 1 1
PA PS
1
0 1
VA 1 LX(1,1) LX(7,2) LX(10,3) LY(1,1) LY(2,2)
VA 0.00 TE(2,2) TE(1,1)
OU SE TV RS SS MI AD=OFF
```

Matriz de correlación análisis factorial confirmatorio (HATCO.COR)

```
  1,000000
 -0,349225    1,000000
  0,509295   -0,487213    1,000000
  0,050414    0,272187   -0,116104    1,000000
  0,611901    0,512981    0,066617    0,298677    1,000000
  0,077115    0,186243   -0,034316    0,788225    0,240808    1,000000
 -0,482631    0,469746   -0,448112    0,199981   -0,055161    0,177294    1,000000
```

Matriz de correlación: Estimación de modelo estructural (STRUCT1.COR)

```
1,0000   0,4112   0,2878   0,3594   0,2683   0,2116   0,2498   0,3046
0,3284   0,2679   0,1423   0,3289   0,5187   0,5100   0,3411
0,4112   1,0000   0,1676   0,1585   0,1413   0,0810   0,0603   0,1270
0,1326   0,0459   0,1043   0,0625   0,0768   0,0896   0,1732
0,2878   0,1676   1,0000   0,7845   0,6763   0,5813   0,6323   0,6903
0,2931   0,1844   0,2890   0,3266   0,4130   0,3297   0,1540
0,3594   0,1585   0,7845   1,0000   0,6370   0,6222   0,6436   0,6667
0,2626   0,1245   0,2492   0,2924   0,3851   0,2718   0,1728
```

0,2683	0,1413	0,6763	0,6370	1,0000	0,6266	0,5378	0,5507
0,3365	0,2296	0,2993	0,1119	0,2647	0,1919	0,1818	
0,2116	0,0810	0,5813	0,6222	0,6266	1,0000	0,6986	0,6251
0,2898	0,2604	0,2194	0,2923	0,3456	0,1514	0,1526	
0,2498	0,0603	0,6323	0,6436	0,5378	0,6986	1,0000	0,6923
0,2065	0,1096	0,1550	0,3258	0,2593	0,2098	0,1070	
0,3046	0,1270	0,6903	0,6667	0,5507	0,6251	0,6923	1,0000
0,1740	0,2536	0,2986	0,2526	0,2622	0,2254	0,1018	
0,3284	0,1326	0,2931	0,2626	0,3365	0,2898	0,2065	0,1740
1,0000	0,3009	0,3071	0,1790	0,3472	0,3479	0,2718	
0,2679	0,0459	0,1844	0,1245	0,2296	0,2604	0,1096	0,2536
0,3009	1,0000	0,6760	0,1137	0,2251	0,2590	0,0522	
0,1423	0,1043	0,2890	0,2492	0,2993	0,2194	0,1550	0,2986
0,3071	0,6760	1,0000	0,1727	0,2048	0,1616	0,0933	
0,3289	0,0625	0,3266	0,2924	0,1119	0,2923	0,3258	0,2526
0,1790	0,1137	0,1727	1,0000	0,4105	0,5550	0,2018	
0,5187	0,0768	0,4130	0,3851	0,2647	0,3456	0,2593	0,2622
0,3472	0,2251	0,2048	0,4105	1,0000	0,5318	0,4112	
0,5100	0,0896	0,3297	0,2718	0,1919	0,1514	0,2098	0,2254
0,3479	0,2590	0,1616	0,5550	0,5318	1,0000	0,4247	
0,3411	0,1732	0,1540	0,1728	0,1818	0,1526	0,1070	0,1018
0,2718	0,0522	0,0933	0,2018	0,4112	0,4247	1,0000	

COMANDOS DE CONTROL PC-MDS

CAPÍTULO 9: ESCALA MULTIDIMENSIONAL

Interpretación de la solución MDS: PROFIT

```
1 10 2 8 0 0 3 0,0                          Tarjetas de control del programa.
(2F9.0)
   0,6077     1,2221
   0,3501     1,3025
  -0,6334     0,9673                        Coordenadas de los estímulos obtenidos del
  -1,1740    -0,9958                          análisis INDSCAL.
  -1,4989    -0,1782
   0,5022    -1,3253
   1,4608    -0,1062
   0,3209    -1,6578
  -1,1906     0,2491
1,2552        0,5224
(10F6,0)
Property 1
  6,94  7,17  7,67  3,22  4,78  5,11  6,56  1,61  8,78  3,17     Evaluación media de cada
Property 2                                                        objeto sobre ocho
  4,00  1,83  6,33  7,67  6,00  5,78  5,50  6,11  7,50  4,17     atributos.
Property 3
  6,94  5,67  3,39  3,67  3,67  6,94  6,44  7,22  4,94  6,11
Property 4
  4,00  3,39  7,33  6,11  7,50  4,22  7,17  4,33  8,22  5,56
Property 5
  5,16  3,47  6,41  5,88  6,06  4,94  5,29  4,82  8,35  4,65
Property 6
  5,11  1,22  5,78  7,89  6,56  3,83  4,28  6,94  8,67  4,72
Property 7
  5,33  3,72  6,33  5,56  6,39  4,72  5,28  5,22  7,33  5,11
Property 8
  4,17  1,56  6,06  8,22  7,72  4,28  3,89  6,33  7,72  5,06
HATCO
Firm B
Firm C
Firm D
Firm E
Firm F                                      Etiquetas.
Firm G
Firm H
Firm I
Firm J
```

Incorporación de preferencias en la solución MDS: PREFMAP

```
10 2 6 0 1 0 2 4 0 1 1 25 0 0 1             Tarjetas de control del programa.
(2F9,0)
   0,6077    1,2221
   0,3501    1,3025
  -0,6334    0,9673
  -1,1740   -0,9958                         Coordenadas de estímulos obtenidos del
  -1,4989   -0,1782                           análisis INDSCAL.
   0,5022   -1,3253
   1,4609   -0,1062
   0,3209   -1,6578
  -1,1906    0,2491
   1,2552    0,5224
(2X,10F6,0)
 1 2 3  5 6 7 4 10 8  1 9
 2 5 2  7 6 9 3  4 1 10 8
```

```
  3  4  1   8  7  6  9   3  5  10   2
  4  4  3  10  2  7  8   6  1   9  5
  6  4  1   8  7  9  3   5  2  10  6
HATCO
Firm B
Firm C
Firm D
Firm E
Firm F
Firm G
Firm H
Firm I
Firm J
SUBJ 1
SUBJ 2
SUBJ 3
SUBJ 4
SUBJ 5
SUBJ 6
```

Valores de preferencia de 10 empresas por 5 encuestados.

Etiquetas.

Desarrollo de mapas perceptuales con datos de tabulación cruzada por el análisis de correspondencias: CORRESP

```
8 10
(10F3,0)
  4   3   1  13   9   6   3  18   2  10
 15  16  15  11  11  14  16  12  14  14
 15  14   6   4   4  15  14  13   7  13
 16  13   8  13   9  17  15  16   6  12
 14  14  10  11  11  14  12  13  10  14
  7  18  13   4   9  16  14   5   4  16
  6   6  14  10  11   8   7   4  14   4
 15  18   9   2   3  15  16   7   8   8
Product Quality
Strategic Orientation
Service
Delivery Speed
Price Level
Salesforce Image
Price Flexibility
Mfgr. Image
HATCO
Firm A
Firm B
Firm C
Firm D
Firm E
Firm F
Firm G
Firm H
Firm I
```

Tarjetas de control de los programas.

Datos de tabulación cruzada de atributos por objetos (empresas).

Etiquetas.

BASE DE DATOS DE HATCO

Base de datos HATCO (X1 a X14)

1	4,1	0,6	6,9	4,7	2,4	2,3	5,2	0	32,0	4,2	1	0	1	1
2	1,8	3,0	6,3	6,6	2,5	4,0	8,4	1	43,0	4,3	0	1	0	1
3	3,4	5,2	5,7	6,0	4,3	2,7	8,2	1	48,0	5,2	0	1	1	2
4	2,7	1,0	7,1	5,9	1,8	2,3	7,8	1	32,0	3,9	0	1	1	1
5	6,0	0,9	9,6	7,8	3,4	4,6	4,5	0	58,0	6,8	1	0	1	3
6	1,9	3,3	7,9	4,8	2,6	1,9	9,7	1	45,0	4,4	0	1	1	2
7	4,6	2,4	9,5	6,6	3,5	4,5	7,6	0	46,0	5,8	1	0	1	1
8	1,3	4,2	6,2	5,1	2,8	2,2	6,9	1	44,0	4,3	0	1	0	2
9	5,5	1,6	9,4	4,7	3,5	3,0	7,6	0	63,0	5,4	1	0	1	3
10	4,0	3,5	6,5	6,0	3,7	3,2	8,7	1	54,0	5,4	0	1	0	2
11	2,4	1,6	8,8	4,8	2,0	2,8	5,8	0	32,0	4,3	1	0	0	1
12	3,9	2,2	9,1	4,6	3,0	2,5	8,3	0	47,0	5,0	1	0	1	2
13	2,8	1,4	8,1	3,8	2,1	1,4	6,6	1	39,0	4,4	0	1	0	1
14	3,7	1,5	8,6	5,7	2,7	3,7	6,7	0	38,0	5,0	1	0	1	1
15	4,7	1,3	9,9	6,7	3,0	2,6	6,8	0	54,0	5,9	1	0	0	3
16	3,4	2,0	9,7	4,7	2,7	1,7	4,8	0	49,0	4,7	1	0	0	3
17	3,2	4,1	5,7	5,1	3,6	2,9	6,2	0	38,0	4,4	1	1	1	2
18	4,9	1,8	7,7	4,3	3,4	1,5	5,9	0	40,0	5,6	1	0	0	2
19	5,3	1,4	9,7	6,1	3,3	3,9	6,8	0	54,0	5,9	1	0	1	3
20	4,7	1,3	9,9	6,7	3,0	2,6	6,8	0	55,0	6,0	1	0	0	3
21	3,3	0,9	8,6	4,0	2,1	1,8	6,3	0	41,0	4,5	1	0	0	2
22	3,4	0,4	8,3	2,5	1,2	1,7	5,2	0	35,0	3,3	1	0	0	1
23	3,0	4,0	9,1	7,1	3,5	3,4	8,4	0	55,0	5,2	1	1	0	3
24	2,4	1,5	6,7	4,8	1,9	2,5	7,2	1	36,0	3,7	0	1	0	1
25	5,1	1,4	8,7	4,8	3,3	2,6	3,8	0	49,0	4,9	1	0	0	2
26	4,6	2,1	7,9	5,8	3,4	2,8	4,7	0	49,0	5,9	1	0	1	3
27	2,4	1,5	6,6	4,8	1,9	2,5	7,2	1	36,0	3,7	0	1	0	1
28	5,2	1,3	9,7	6,1	3,2	3,9	6,7	0	54,0	5,8	1	0	1	3
29	3,5	2,8	9,9	3,5	3,1	1,7	5,4	0	49,0	5,4	1	0	1	3
30	4,1	3,7	5,9	5,5	3,9	3,0	8,4	1	46,0	5,1	0	1	0	2
31	3,0	3,2	6,0	5,3	3,1	3,0	8,0	1	43,0	3,3	0	1	0	1
32	2,8	3,8	8,9	6,9	3,3	3,2	8,2	0	53,0	5,0	1	1	0	3
33	5,2	2,0	9,3	5,9	3,7	2,4	4,6	0	60,0	6,1	1	0	0	3
34	3,4	3,7	6,4	5,7	3,5	3,4	8,4	1	47,3	3,8	0	1	0	1
35	2,4	1,0	7,7	3,4	1,7	1,1	6,2	1	35,0	4,1	0	1	0	1
36	1,8	3,3	7,5	4,5	2,5	2,4	7,6	1	39,0	3,6	0	1	1	1
37	3,6	4,0	5,8	5,8	3,7	2,5	9,3	1	44,0	4,8	0	1	1	2
38	4,0	0,9	9,1	5,4	2,4	2,6	7,3	0	46,0	5,1	1	0	1	3
39	0,0	2,1	6,9	5,4	1,1	2,6	8,9	1	29,0	3,9	0	1	1	1
40	2,4	2,0	6,4	4,5	2,1	2,2	8,8	1	28,0	3,3	0	1	1	1
41	1,9	3,4	7,6	4,6	2,6	2,5	7,7	1	40,0	3,7	0	1	1	1
42	5,9	0,9	9,6	7,8	3,4	4,6	4,5	0	58,0	6,7	1	0	1	3
43	4,9	2,3	9,3	4,5	3,6	1,3	6,2	0	53,0	5,9	1	0	0	3
44	5,0	1,3	8,6	4,7	3,1	2,5	3,7	0	48,0	4,8	1	0	0	2
45	2,0	2,6	6,5	3,7	2,4	1,7	8,5	1	38,0	3,2	0	1	1	1
46	5,0	2,5	9,4	4,6	3,7	1,4	6,3	0	54,0	6,0	1	0	0	3
47	3,1	1,9	10,0	4,5	2,6	3,2	3,8	0	55,0	4,9	1	0	1	3
48	3,4	3,9	5,6	5,6	3,6	2,3	9,1	1	43,0	4,7	0	1	1	2
49	5,8	0,2	8,8	4,5	3,0	2,4	6,7	0	57,0	4,9	1	0	1	3
50	5,4	2,1	8,0	3,0	3,8	1,4	5,2	0	53,0	3,8	1	0	1	3
51	3,7	0,7	8,2	6,0	2,1	2,5	5,2	0	41,0	5,0	1	0	0	2
52	2,6	4,8	8,2	5,0	3,6	2,5	9,0	1	53,0	5,2	0	1	1	2
53	4,5	4,1	6,3	5,9	4,3	3,4	8,8	1	50,0	5,5	0	1	0	2
54	2,8	2,4	6,7	4,9	2,5	2,6	9,2	1	32,0	3,7	0	1	1	1
55	3,8	0,8	8,7	2,9	1,6	2,1	5,6	0	39,0	3,7	1	0	0	1
56	2,9	2,6	7,7	7,0	2,8	3,6	7,7	0	47,0	4,2	1	1	1	2
57	4,9	4,4	7,4	6,9	4,6	4,0	9,6	1	62,0	6,2	0	1	0	2
58	5,4	2,5	9,6	5,5	4,0	3,0	7,7	0	65,0	6,0	1	0	0	3
59	4,3	1,8	7,6	5,4	3,1	2,5	4,4	0	46,0	5,6	1	0	1	3

60	2,3	4,5	8,0	4,7	3,3	2,2	8,7	1	50,0	5,0	0	1	1	2
61	3,1	1,9	9,9	4,5	2,6	3,1	3,8	0	54,0	4,8	1	0	1	3
62	5,1	1,9	9,2	5,8	3,6	2,3	4,5	0	60,0	6,1	1	0	0	3
63	4,1	1,1	9,3	5,5	2,5	2,7	7,4	0	47,0	5,3	1	0	1	3
64	3,0	3,8	5,5	4,9	3,4	2,6	6,0	0	36,0	4,2	1	1	1	2
65	1,1	2,0	7,2	4,7	1,6	3,2	10,0	1	40,0	3,4	0	1	1	1
66	3,7	1,4	9,0	4,5	2,6	2,3	6,8	0	45,0	4,9	1	0	0	2
67	4,2	2,5	9,2	6,2	3,3	3,9	7,3	0	59,0	6,0	1	0	0	3
68	1,6	4,5	6,4	5,3	3,0	2,5	7,1	1	46,0	4,5	0	1	0	2
69	5,3	1,7	8,5	3,7	3,5	1,9	4,8	0	58,0	4,3	1	0	0	3
70	2,3	3,7	8,3	5,2	3,0	2,3	9,1	1	49,0	4,8	0	1	1	2
71	3,6	5,4	5,9	6,2	4,5	2,9	8,4	1	50,0	5,4	0	1	1	2
72	5,6	2,2	8,2	3,1	4,0	1,6	5,3	0	55,0	3,9	1	0	1	3
73	3,6	2,2	9,9	4,8	2,9	1,9	4,9	0	51,0	4,9	1	0	0	3
74	5,2	1,3	9,1	4,5	3,3	2,7	7,3	0	60,0	5,1	1	0	1	3
75	3,0	2,0	6,6	6,6	2,4	2,7	8,2	1	41,0	4,1	0	1	0	1
76	4,2	2,4	9,4	4,9	3,2	2,7	8,5	0	49,0	5,2	1	0	1	2
77	3,8	0,8	8,3	6,1	2,2	2,6	5,3	0	42,0	5,1	1	0	0	2
78	3,3	2,6	9,7	3,3	2,9	1,5	5,2	0	47,0	5,1	1	0	1	3
79	1,0	1,9	7,1	4,5	1,5	3,1	9,9	1	39,0	3,3	0	1	1	1
80	4,5	1,6	8,7	4,6	3,1	2,1	6,8	0	56,0	5,1	1	0	0	3
81	5,5	1,8	8,7	3,8	3,6	2,1	4,9	0	59,0	4,5	1	0	0	3
82	3,4	4,6	5,5	8,2	4,0	4,4	6,3	0	47,3	5,6	1	1	1	2
83	1,6	2,8	6,1	6,4	2,3	3,8	8,2	1	41,0	4,1	0	1	0	1
84	2,3	3,7	7,6	5,0	3,0	2,5	7,4	0	37,0	4,4	1	1	0	1
85	2,6	3,0	8,5	6,0	2,8	2,8	6,8	1	53,0	5,6	0	1	0	2
86	2,5	3,1	7,0	4,2	2,8	2,2	9,0	1	43,0	3,7	0	1	1	1
87	2,4	2,9	8,4	5,9	2,7	2,7	6,7	1	51,0	5,5	0	1	0	2
88	2,1	3,5	7,4	4,8	2,8	2,3	7,2	0	36,0	4,3	1	1	0	1
89	2,9	1,2	7,3	6,1	2,0	2,5	8,0	1	34,0	4,0	0	1	1	1
90	4,3	2,5	9,3	6,3	3,4	4,0	7,4	0	60,0	6,1	1	0	0	3
91	3,0	2,8	7,8	7,1	3,0	3,8	7,9	0	49,0	4,4	1	1	1	2
92	4,8	1,7	7,6	4,2	3,3	1,4	5,8	0	39,0	5,5	1	0	0	2
93	3,1	4,2	5,1	7,8	3,6	4,0	5,9	0	43,0	5,2	1	1	1	2
94	1,9	2,7	5,0	4,9	2,2	2,5	8,2	1	36,0	3,6	0	1	0	1
95	4,0	0,5	6,7	4,5	2,2	2,1	5,0	0	31,0	4,0	1	0	1	1
96	0,6	1,6	6,4	5,0	0,7	2,1	8,4	1	25,0	3,4	0	1	1	1
97	6,1	0,5	9,2	4,8	3,3	2,8	7,1	0	60,0	5,2	1	0	1	3
98	2,0	2,8	5,2	5,0	2,4	2,7	8,4	1	38,0	3,7	0	1	0	1
99	3,1	2,2	6,7	6,8	2,6	2,9	8,4	1	42,0	4,3	0	1	0	1
100	2,5	1,8	9,0	5,0	2,2	3,0	6,0	0	33,0	4,4	1	0	0	1

Pre-test de las bases de datos de HATCO (X_1 a X_{14}) con datos ausentes

(Indicado con a , Leáse con mismo formato en la base de datos HATCO)

201	3,3	0,9	8,6	4,0	2,1	1,8	6,3	0	41,0	4,5	1	0	0	2
202	.	0,4	.	2,5	1,2	1,7	5,2	0	35,0	3,3	1	0	0	1
203	3,0	.	9,1	7,1	3,5	3,4	.	0	55,0	5,2	1	1	0	3
204	.	1,5	.	4,8	1,9	2,5	7,2	1	36,0	.	0	1	0	1
205	5,1	1,4	.	4,8	3,3	2,6	3,8	0	49,0	4,9	1	0	0	2
206	4,6	2,1	7,9	5,8	3,4	2,8	4,7	0	49,0	5,9	1	0	1	3
207	.	1,5	.	4,8	1,9	2,5	7,2	1	36,0	.	0	1	0	1
208	5,2	1,3	9,7	6,1	3,2	3,9	6,7	0	54,3	5,8	1	0	1	3
209	3,5	2,8	9,9	3,5	3,1	1,7	5,4	0	49,0	5,4	1	0	1	3
210	4,1	3,7	5,9	.	.	.	.	.	.	.	0	1	0	2
211	3,0	2,8	7,8	7,1	3,0	3,8	7,9	0	49,0	4,4	1	1	1	2
212	4,8	1,7	7,6	4,2	3,3	1,4	5,8	0	39,0	5,5	1	0	0	2
213	3,1	.	0,	7,8	3,6	4,0	5,9	0	43,0	5,2	1	1	1	2
214	.	2,7	5,0	.	2,2	.	.	1	.	3,6	.	1	.	1
215	4,0	0,5	6,7	4,5	2,2	2,1	5,0	.	31,0	4,0	1	0	1	1
216	.	1,6	6,4	5,0	.	2,1	8,4	1	25,0	3,4	0	1	1	1
217	6,1	0,5	9,2	4,8	3,3	2,8	7,1	0	60,0	5,2	1	0	1	3
218	.	2,8	5,2	5,0	.	2,7	8,4	1	38,0	3,7	0	1	0	1

219	3,1	2,2	6,7	6,8	2,6	2,9	.	1	.	4,3	0	1	0	1
220	6,5	.	9,0	7,0	3,2	3,7	8,0	0	33,0	5,4	1	0	0	1
221	.	1,6	.	4,8	2,0	2,8	.	0	32,0	4,3	1	0	0	1
222	3,9	2,2	.	4,6	.	2,5	8,3	0	47,0	5,0	1	0	1	2
223	2,8	1,4	8,1	3,8	2,1	1,4	6,6	1	39,0	4,4	0	1	0	1
224	.	.	8,6	5,7	2,7	3,7	6,7	0	.	5,0	1	0	1	1
225	4,7	1,3	.	0,	3,0	2,6	6,8	0	54,0	5,9	1	0	0	1
226	3,4	2,0	9,7	4,7	2,7	1,7	4,8	0	49,0	4,7	1	0	0	3
227	3,2	.	5,7	5,1	3,6	2,9	6,2	0	.	4,4	1	1	1	2
228	.	1,8	7,7	.	3,4	1,5	5,9	0	40,0	5,6	1	0	0	2
229	5,3	1,4	9,7	6,1	.	3,9	6,8	0	54,0	5,9	1	0	1	3
230	4,7	1,3	9,9	6,7	3,0	2,6	6,8	0	55,0	6,0	1	0	0	3
231	3,7	0,7	8,2	6,0	2,1	2,5	.	0	41,0	5,0	1	0	0	2
232	.	.	8,2	5,0	3,6	2,5	9,0	1	53,0	5,2	0	1	1	2
233	4,5	.	.	5,9	.	.	8,8	1	50,0	.	0	.	0	.
234	2,8	2,4	6,7	4,9	2,5	2,6	9,2	1	32,0	3,7	0	1	1	1
235	3,8	0,8	8,7	2,9	1,6	.	5,6	0	39,0	.	1	0	0	1
236	2,9	2,6	7,7	7,0	2,8	3,6	7,7	0	47,0	4,2	1	1	1	2
237	4,9	.	7,4	6,9	4,6	4,0	9,6	1	62,0	6,2	0	1	0	2
238	.	2,5	9,6	5,5	4,0	3,0	7,7	0	65,0	6,0	1	0	0	3
239	4,3	1,8	7,6	5,4	3,1	2,5	4,4	0	46,0	5,6	1	0	1	3
240	.	1,5	9,9	2,7	1,3	1,2	1,7	1	50,0	5,0	0	1	1	2
241	3,1	1,9	.	4,5	.	3,1	3,8	0	54,0	4,8	1	0	1	3
242	5,1	1,9	9,2	5,8	3,6	2,3	4,5	0	60,0	6,1	1	0	0	3
243	4,1	1,1	9,3	5,5	2,5	2,7	7,4	0	47,0	5,3	1	0	1	3
244	3,0	3,8	5,5	4,9	3,4	2,6	6,0	0	.	4,2	1	1	1	2
245	.	2,0	.	4,7	.	3,2	.	1	.	3,4	0	.	1	0,
246	3,7	1,4	9,0	.	2,6	2,3	6,8	0	45,0	4,9	1	0	0	2
247	4,2	2,5	9,2	6,2	3,3	3,9	7,3	0	59,0	6,0	1	0	0	3
248	.	0,	6,4	5,3	3,0	2,5	7,1	1	46,0	4,5	0	1	0	2
249	5,3	.	8,5	3,7	3,5	1,9	4,8	0	58,0	4,3	1	0	0	3
250	.	3,7	.	5,2	3,0	2,3	9,1	1	49,0	4,8	0	1	1	2
251	3,0	3,2	6,0	5,3	3,1	3,0	8,0	1	43,0	3,3	0	1	0	1
252	2,8	3,8	8,9	6,9	3,3	3,2	8,2	0	53,0	5,0	1	1	0	3
253	.	2,0	9,3	5,9	3,7	2,4	4,6	0	60,0	6,1	1	0	0	3
254	3,4	3,7	6,4	5,7	3,5	3,4	8,4	1	47,0	3,8	0	1	0	1
255	.	1,0	.	3,4	1,7	1,1	6,2	1	35,0	4,1	0	1	0	1
256	.	3,3	7,5	4,5	2,5	2,4	7,6	1	39,0	3,6	0	1	1	1
257	3,6	.	0,	5,8	3,7	2,5	9,3	1	44,0	4,8	0	1	1	2
258	4,0	0,9	9,1	5,4	2,4	2,6	7,3	0	46,0	5,1	1	0	1	3
259	.	2,1	6,9	5,4	1,1	2,6	8,9	1	29,0	3,9	0	1	1	1
260	.	2,0	6,4	4,5	2,1	2,2	8,8	1	28,0	3,3	0	1	1	1
261	3,6	.	.	6,2	4,5	.	.	1	.	0,	.	1	1	2
262	5,6	2,20	8,2	3,1	4,0	1,6	5,3	0	55,0	3,9	1	0	1	3
263	3,6	.	9,9	.	.	.	.	.	.	4,9	1	0	0	3
264	5,2	1,3	9,1	4,5	3,3	2,7	7,3	0	60,0	5,1	1	0	1	3
265	3,0	2,0	6,6	6,6	2,4	2,7	8,2	1	41,0	4,1	0	1	0	1
266	4,2	2,4	9,4	4,9	3,2	2,7	8,5	0	49,0	5,2	1	0	1	2
267	3,8	0,8	.	0,	2,2	2,6	5,3	0	42,0	5,1	1	0	0	2
268	3,3	2,6	9,70	3,30	2,9	1,5	5,2	0	47,0	.	1	0	1	3
269	.	1,9	.	4,5	1,5	3,1	9,9	1	39,0	3,3	0	1	1	1
270	4,5	1,6	8,7	4,6	3,1	2,1	6,8	0	56,0	5,1	1	0	0	3

Serie de datos del análisis conjunto: Evaluaciones de 22 estímulos por 100 encuestados

104	4	6	5	4	4	6	4	4	4	4	4	4	4	5	5	4	4	4	4	4	6	6
107	6	3	5	2	3	1	1	6	6	6	7	4	1	6	6	6	6	1	7	7	1	7
109	7	7	6	7	3	3	3	6	3	5	2	3	2	2	6	3	3	3	6	3	3	3
110	5	5	7	7	1	5	1	1	4	5	5	5	1	5	5	2	6	1	3	5	1	5
120	7	7	1	6	1	1	1	1	1	5	5	1	1	1	7	1	1	1	5	6	1	1
123	7	7	5	5	1	1	1	2	2	1	7	1	1	2	7	7	7	1	7	7	1	7
129	7	7	7	7	7	7	7	7	6	7	6	7	6	7	7	6	6	7	7	6	7	6
133	5	7	2	7	2	2	2	3	3	2	5	2	1	6	5	2	2	2	2	6	2	7

135	7	7	2	5	1	1	1	2	2	2	7	1	1	7	4	5	2	1	2	5	2	5
144	6	7	3	6	6	2	3	3	5	5	7	7	2	2	7	3	6	5	5	5	5	3
150	7	7	5	5	3	6	3	4	7	7	5	2	5	7	6	7	4	3	6	7	3	6
155	7	7	3	6	2	1	1	5	1	5	1	3	2	2	7	1	1	1	5	1	1	1
156	7	6	5	3	5	5	2	6	5	2	7	2	5	6	7	5	5	6	6	7	2	6
161	6	6	5	5	1	1	1	3	2	5	2	1	1	5	7	3	1	1	5	2	1	2
162	7	6	3	6	3	5	3	6	3	2	6	6	2	5	7	4	2	2	4	7	6	5
167	4	5	3	3	2	3	3	2	6	5	6	3	3	3	4	2	3	5	3	6	5	5
168	7	7	3	2	2	2	1	1	2	3	7	1	1	7	7	7	3	2	3	7	2	7
170	6	6	4	2	1	3	2	1	2	5	7	3	3	1	7	2	6	2	4	7	2	5
173	6	6	4	4	3	3	3	5	2	4	6	3	3	5	5	5	4	3	4	6	5	6
174	6	5	6	6	6	1	3	6	1	6	1	5	1	6	7	1	1	5	6	1	3	1
178	7	7	2	5	1	1	1	1	1	5	1	1	1	1	6	1	1	1	5	3	2	1
180	7	7	2	6	6	1	2	2	2	5	2	7	1	7	7	2	1	5	7	2	7	2
181	6	6	3	5	4	2	2	3	3	5	6	4	2	3	6	4	6	2	6	5	4	6
187	6	7	6	7	2	2	2	5	5	6	6	3	1	4	7	5	2	3	6	7	3	6
190	6	6	5	5	1	1	1	3	5	5	7	1	1	3	6	3	5	1	5	5	1	6
192	7	7	5	6	2	1	2	2	2	2	2	2	1	4	5	1	1	2	5	3	3	3
193	7	7	6	6	2	1	6	6	1	5	4	5	2	3	7	1	2	2	7	2	7	2
194	2	6	6	2	6	2	2	4	2	4	2	4	3	4	6	1	4	6	4	2	6	2
195	5	5	5	5	5	4	5	5	4	5	4	5	4	5	5	4	4	4	4	4	5	4
197	6	7	5	6	3	3	4	6	4	4	6	3	7	4	4	4	5	3	3	7	6	6
198	7	7	5	7	4	6	5	7	5	4	2	7	2	4	7	1	1	2	5	2	6	2
200	5	4	2	2	3	5	5	4	5	3	4	2	2	2	3	1	2	3	5	4	3	1
209	6	6	1	2	2	2	1	2	5	2	5	1	1	2	6	1	1	1	2	7	1	4
211	5	7	5	7	2	2	1	5	5	2	5	1	2	2	7	5	2	2	1	7	7	7
222	7	7	3	3	2	2	5	5	3	6	5	2	2	6	7	3	4	2	5	5	2	5
225	7	7	4	6	2	2	2	6	2	4	3	3	1	4	5	2	2	3	5	3	4	2
229	7	7	7	7	1	4	5	3	6	2	6	3	1	2	2	7	1	4	1	1	5	1
231	5	5	2	2	1	1	1	5	1	2	6	2	1	3	5	2	2	1	6	4	1	3
233	6	5	2	1	7	5	6	2	1	7	5	3	3	3	7	6	2	1	4	4	1	4
234	5	5	4	5	4	4	5	5	4	4	4	4	4	4	5	4	4	5	6	4	5	4
235	5	7	6	6	2	3	6	3	3	6	7	6	3	7	7	6	6	3	5	5	3	5
236	4	6	2	4	1	1	1	2	2	2	4	1	1	5	7	2	1	1	2	2	1	2
240	2	2	5	7	2	6	2	2	4	2	2	2	2	2	7	6	1	5	4	6	6	4
246	7	3	6	4	2	6	5	6	5	5	5	4	3	3	2	2	3	3	1	1	4	1
249	4	4	6	7	2	1	2	5	5	4	3	2	1	4	6	4	2	2	6	4	2	3
250	7	4	3	3	4	5	3	3	3	4	7	3	5	5	4	5	4	4	3	4	5	7
251	4	4	1	4	1	1	1	2	2	2	2	1	1	4	3	2	1	1	1	3	5	2
254	5	7	2	1	1	1	1	3	1	3	3	1	1	2	4	1	1	1	3	1	1	1
258	7	1	1	3	1	2	1	1	3	1	7	1	3	1	7	1	1	3	4	4	1	4
260	5	5	5	5	5	2	4	5	2	5	2	6	2	5	5	1	2	2	5	2	2	2
261	6	6	4	6	2	2	2	4	6	6	6	2	2	4	6	4	4	2	4	4	2	4
266	6	7	7	5	5	6	4	2	3	3	3	2	2	1	4	3	4	2	1	2	1	1
271	7	5	4	3	3	2	3	4	5	6	6	3	4	4	2	2	2	1	2	1	1	1
272	7	7	6	5	6	6	3	3	4	5	5	6	2	2	3	4	1	4	1	3	3	1
277	6	6	5	5	4	3	4	3	3	3	5	4	1	2	3	1	2	2	2	1	1	1
285	5	4	2	3	5	5	4	3	4	2	4	4	3	3	1	1	5	2	2	1	2	1
287	3	5	2	2	1	1	1	1	1	1	3	1	1	2	2	3	1	1	1	5	1	2
288	7	4	4	3	4	4	3	5	6	6	2	3	3	3	2	2	1	2	3	3	1	1
289	5	4	2	4	4	5	5	2	4	2	4	5	3	3	3	2	3	1	3	3	2	2
300	6	5	4	5	2	2	2	4	4	5	5	2	2	5	6	3	5	2	5	5	2	5
302	6	6	6	7	5	3	4	5	3	5	3	5	3	5	6	2	2	4	7	3	7	3
303	6	6	5	7	4	2	2	3	3	3	3	6	2	3	7	2	3	5	6	3	7	2
306	5	5	6	7	5	2	4	4	2	5	1	4	2	6	7	3	3	5	7	3	7	2
308	7	7	3	6	3	3	3	3	3	3	6	3	6	6	6	7	3	3	3	5	7	5
309	7	6	5	7	3	1	1	3	3	3	1	2	2	6	6	3	2	2	4	1	4	3
310	6	7	3	7	3	3	3	4	4	4	4	5	2	5	7	2	2	7	5	2	6	3
317	7	7	6	6	3	3	3	5	3	5	3	4	3	5	5	3	3	5	5	3	2	3
318	3	4	5	6	5	2	5	6	1	6	2	5	3	6	7	4	2	5	6	4	6	4
319	1	5	6	5	2	1	4	3	2	6	2	5	5	4	6	4	6	4	6	4	6	4
323	4	4	1	6	1	6	1	4	1	4	4	1	4	6	6	6	4	4	1	4	1	4
324	6	7	7	6	6	2	5	6	2	5	5	7	3	6	7	1	2	7	7	2	7	5
325	7	7	5	5	5	7	6	5	4	3	3	3	4	6	7	7	5	7	6	5	7	7

330	4	5	5	6	3	6	3	4	5	4	6	5	4	4	6	5	4	7	5	5	6	4
336	3	1	4	1	4	4	5	5	5	5	3	2	5	7	4	6	5	4	2	3	1	4
339	6	6	2	4	1	1	1	2	2	6	2	1	1	6	7	1	2	1	3	4	1	3
348	3	3	1	1	1	1	1	1	1	1	3	1	1	3	5	6	1	1	1	6	1	6
350	6	6	3	7	5	2	3	5	2	5	2	5	2	5	5	2	2	5	6	3	6	3
352	6	7	4	4	1	1	1	4	4	4	6	1	1	6	7	1	1	1	4	1	1	1
353	7	7	5	5	3	3	3	5	1	3	1	1	1	7	7	3	3	3	5	1	1	7
354	5	5	4	6	1	1	2	4	3	5	2	2	6	6	6	6	5	2	6	6	5	6
356	4	5	2	4	1	1	1	4	3	2	5	2	1	3	5	7	2	1	3	6	2	6
363	7	6	4	6	1	1	1	5	3	6	7	2	6	5	7	6	5	2	6	5	5	6
366	5	5	6	3	3	3	3	4	4	3	4	2	2	5	6	4	2	1	6	4	3	3
368	7	7	6	6	3	1	2	4	2	5	3	2	1	3	6	3	2	1	3	2	2	2
370	6	3	5	5	3	2	4	6	5	5	2	4	4	6	7	5	4	4	6	3	2	3
372	7	6	7	5	5	3	5	5	2	6	2	5	3	6	7	2	3	4	7	2	3	2
381	1	3	3	2	2	1	7	2	2	4	4	3	3	2	2	1	2	2	3	3	4	2
382	3	3	3	2	2	1	2	4	7	5	2	2	1	2	3	3	3	4	4	6	1	1
385	4	7	7	5	7	5	7	5	6	4	7	1	4	3	4	4	4	3	1	2	1	1
394	6	6	4	5	1	2	3	7	6	7	1	1	1	6	2	3	6	4	5	4	5	5
396	7	7	5	5	1	1	1	5	5	5	7	1	1	7	7	7	5	1	5	7	1	7
399	4	5	2	1	1	1	1	1	2	2	4	1	1	4	5	2	1	1	1	4	1	2
401	6	6	3	3	1	1	1	3	1	3	5	1	1	5	7	6	1	1	4	5	1	5
403	5	5	6	5	1	1	1	3	2	5	2	2	1	0	5	4	2	1	5	4	5	2
412	7	7	6	6	1	1	1	5	1	6	1	1	1	7	6	1	1	1	5	1	1	1
413	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1
414	6	6	5	7	3	4	3	5	2	5	2	2	3	4	6	2	1	2	5	4	3	2
416	5	7	2	6	2	2	1	4	3	1	6	3	2	1	7	2	1	1	4	5	2	2
417	5	5	2	5	1	1	2	2	2	2	1	1	1	4	6	3	2	1	4	4	2	3
418	6	7	2	7	1	4	1	3	6	4	7	1	5	2	1	2	4	2	3	7	5	7

Datos de preferencia: Valores de similitud para 18 encuestados

Encuestados 1-5	Encuestados 6-10	Encuestados 11-15	Encuestados 16-18
4	8	3	7
6 9	7 4	3 6	7 5
6 6 1	1 1 2	1 1 1	1 1 2
6 6 4 8	1 1 2 8	1 1 2 4	1 1 2 5
3 2 2 3 3	2 1 1 1 1	3 1 1 2 1	5 1 1 2 1
3 5 2 2 2 4	1 2 2 1 1 1	1 1 4 2 1 2	1 2 5 1 1 2
2 2 4 4 4 2 4	1 1 2 1 1 1 1	1 1 1 1 1 2 1	1 1 2 2 2 1 1
4 2 9 2 8 8 2 2	1 6 8 1 8 6 1 1	1 1 7 2 7 6 3 2	1 7 2 1 7 7 1 1
8 9 4 2 2 5 9 3 2	7 9 1 1 1 1 6 1 1	5 8 2 2 2 2 7 2 1	7 8 1 1 1 2 7 1 1
8	6	8	4
7 6	4 3	8 6	4 5
1 1 3	1 1 3	6 5 8	2 2 2
1 1 2 8	1 1 3 6	6 6 6 8	2 2 4 8
3 1 1 3 1	4 1 1 4 1	6 6 6 8 6	4 2 2 4 2
1 2 3 1 1 3	2 4 4 1 1 4	6 5 6 5 5 8	2 3 4 2 2 2
1 1 2 1 1 2 1	1 1 4 2 1 3 1	6 6 6 7 6 6 6	2 2 4 2 2 4 2
1 4 1 1 7 1 1 1	2 3 7 2 7 4 1 2	6 5 8 6 8 7 4 5	2 4 6 4 5 4 3 3
1 6 2 1 1 2 1 1 1	8 6 1 2 2 2 6 2 2	7 7 5 6 6 6 8 5 5	5 5 3 3 3 5 6 3 3
9	8	9	6
8 7	7 2	8 6	7 9
5 5 7	1 1 2	1 1 5	1 1 8
6 6 7 8	1 1 2 7	3 3 6 8	1 1 6 7
7 5 5 7 6	2 4 1 1 1	6 2 2 5 3	6 1 1 8 1
6 7 7 6 6 7	1 1 1 1 1 2	3 6 6 1 1 5	1 6 6 1 1 8
6 6 8 6 6 7 5	1 1 2 1 1 2 1	3 3 6 4 4 7 3	1 1 7 1 1 7 1
5 8 9 7 8 8 7 6	1 6 6 1 2 2 1 1	3 8 8 2 6 5 5 6	1 7 9 1 9 9 2 1
8 8 7 7 7 8 9 7 7	2 2 1 1 1 4 6 1 1	6 6 3 4 4 2 8 2 2	9 9 1 1 1 6 9 1 1
7	3	7	
6 4	3 4	4 4	
3 3 5	2 2 2	2 1 1	

Encuestados 1-5	Encuestados 6-10	Encuestados 11-15	Encuestados 16-18
2 2 3 8	2 2 2 6	1 1 3 5	
3 3 3 4 3	2 2 2 4 2	6 1 1 1 1	
3 5 4 3 3 4	2 3 3 3 3 3	1 2 5 1 1 1	
3 3 5 3 4 5 3	3 3 3 3 3 2 2	1 1 2 1 1 1 1	
3 4 8 4 8 4 3 3	2 5 7 2 8 7 4 2	1 3 3 1 7 4 1 1	
7 6 3 3 3 4 8 2 3	7 8 5 2 2 4 8 2 2	4 7 1 1 1 1 5 1 1	
5	9	8	
2 7	9 3	7 7	
2 2 1	1 7 2	5 6 7	
3 2 8 5	2 2 7 9	6 7 5 7	
3 3 2 2 2	2 2 2 7 2	6 5 5 7 6	
2 2 2 2 2 4	2 2 2 2 2 7	7 5 6 5 5 6	
2 2 2 2 2 1 2	2 2 2 2 2 7 3	5 6 5 5 5 6 5	
2 7 7 1 6 5 2 1	3 2 8 2 8 7 2 2	5 7 6 5 8 7 6 6	
5 5 2 1 1 1 6 2 2	8 8 2 2 2 2 8 2 2	7 8 6 5 5 6 8 6 6	

Glosario

Adición avanzada (*Fordward)* Método que selecciona las variables para su inclusión en el modelo de regresión empezando sin variables en el modelo y después añadiendo variables basadas en su contribución a la predicción.

Alfa (α) Véase *error de Tipo I.*

Alfa de Cronbach Media de fiabilidad habitualmente utilizada respecto a un conjunto de dos o más indicadores de un *constructo*. Los valores van desde 0 a 1. Los valores desde 0,60 hasta 0,70 se consideran el límite inferior de aceptabilidad.

Algoritmo Conjunto de reglas de procedimiento; parecido a una ecuación.

Algoritmo genético Modelos de aprendizaje desarrollados sobre los principios de la evolución. Copian el proceso evolutivo mediante el uso de la selección natural, empezando con una serie de soluciones alternativas y seleccionando algunas de ellas en sucesivos procesos hasta llegar a conseguir una solución definitiva.

Almacenamiento de datos Asimilación e integración de los datos desde fuentes internas y externas en un formato adecuado para las aplicaciones de *procesamiento analítico en línea* y *sistemas de apoyo a la decisión.*

Ampliación de la muestra Grupo de objetos que no se consideran del total de la muestra cuando se calcula la(s) función(es) discriminante(s) o el modelo de regresión logística. Este grupo se emplea posteriormente para validar la función discriminante o el modelo de regresión logística sobre una muestra separada de encuestados. También denominada muestra de validación.

Análisis agregado Aproximación al análisis multidimensional en el cual el *mapa perceptual* está generado por un grupo de evaluaciones de encuestados sobre *objetos*. Este mapa perceptual compuesto puede crearse con un programa informático o por el investigador para encontrar unos pocos sujetos «medios» o representativos.

Análisis cluster Técnica multivariante cuyo objetivo es agrupar a los encuestados o los casos con perfiles similares sobre una serie de características definidas. Similar al *análisis factorial Q.*

Análisis confirmatorio Uso de una técnica multivariante para contrastar (confirmar) una relación preespecificada. Por ejemplo, supóngase la hipótesis de que sólo dos variables deberían ser los predictores de una variable dependiente. Si contrastamos empíricamente la significación de esos dos predictores y la no significación del resto, este contraste es un análisis confirmatorio. Es el opuesto a un *análisis exploratorio.*

Análisis conjunto adaptativo Metodología del análisis conjunto que descansa en la información aportada por los encuestados (por ejemplo, importancia de los atributos) para adaptar el diseño conjunto y hacer el problema más sencillo. Ejemplos de este tipo de técnica son los *modelos híbridos* o *adaptativos* y *autoexplicados.*

Análisis conjunto tradicional Metodología que emplea los principios «clásicos» del análisis conjunto, utilizando un modelo *aditivo* de preferencias del consumidor y comparaciones *pareadas* o métodos de presentación de *perfil completo.*

Análisis de componentes principales Modelo factorial en el que los factores se basan en la varianza total. En el análisis de componentes principales, se usan las unidades que aparecen en la diagonal de la *matriz de correlación;* este procedimiento implica, por lo que se refiere al cálculo, que toda la varianza es *común* o compartida.

Análisis de correspondencias *Aproximación de composición* a la elaboración de mapas perceptuales que relacionan categorías de una *tabla de contingencia.* La mayoría de las aplicaciones contienen un conjunto de *objetos* y atributos, con los resultados representando tanto los objetos como los atributos en un *mapa perceptual.*

Análisis de correspondencias múltiples Forma de *análisis de correspondencia* que contiene tres o más variables categóricas relacionadas en un espacio perceptual común.

Análisis de dependencia Grupo de técnicas estadísticas caracterizadas por considerar una variable o un conjunto de ellas como *dependientes* y el resto como *independientes.* El objetivo es predecir la(s) variable(s) dependiente(s) a partir de la(s) independiente(s). Un ejemplo lo constituye el análisis de regresión.

Análisis de interdependencia Grupo de técnicas estadísticas en que las variables no se dividen en *dependientes* e *independientes* (por ejemplo, análisis factorial), sino que se consideran como un único conjunto.

Análisis de la varianza (ANOVA) Técnica estadística empleada para determinar si las muestras provienen de poblaciones con medias iguales. El análisis univariante de la varianza utiliza una variable dependiente, mientras que el análisis multivariante de la varianza compara muestras basadas en dos o más variables dependientes.

Análisis de reducción Contraste para detectar si el poder discriminante de una variable dependiente aumenta después de que ha sido tenido en cuenta el efecto de otras variables dependientes. Similar al análisis discriminante o a la regresión por etapas, este procedimiento, considerando un orden de entrada específico, determina cuánto aporta una variable dependiente adicional a la explicación de las diferencias entre los grupos en el análisis MANOVA.

Análisis de relaciones Método que emplea correlaciones bivariantes simples para estimar las relaciones en un sistema de ecuaciones estructurales. El método se basa en la especificación de relaciones en una serie de ecuaciones de regresión (representada gráficamente en un *diagrama de secuencias*) que pueden ser estimadas a continuación mediante la determinación de la cantidad de correlaciones atribuibles a cada efecto en cada ecuación simultáneamente. Cuando se emplea con relaciones múltiples entre *constructos latentes* y un *modelo de medida*, se le denomina *modelo de ecuaciones estructurales*.

Análisis desagregado Aproximación al análisis multidimensional en el que el investigador genera *mapas perceptuales* encuestado por encuestado. Los resultados pueden ser difíciles de generalizar para todos los encuestados. Por tanto, el investigador puede intentar crear menos mapas mediante algún proceso de *análisis agregado*, en el cual se encuentren combinados los resultados de los encuestados.

Análisis exploratorio Análisis que define posibles relaciones sólo de la forma más general y a continuación deja a las técnicas multivariantes la estimación de la(s) relación(es). Opuesto al *análisis confirmatorio*, el investigador no busca «confirmar» cualquier estimación especificada antes del análisis, sino que deja al método y a los datos definir la naturaleza de las relaciones. Un ejemplo de este tipo de análisis es la regresión múltiple por etapas, en el que el método consiste en añadir variables predictor hasta que se cumpla algún criterio.

Análisis factorial común Modelo factorial en el que los factores se basan en una matriz de correlación reducida. Es decir, se insertan las *varianzas compartidas* en el diagonal de la matriz *de correlación,* y los factores extraídos se basan solamente en la *varianza común,* con la exclusión de las *varianzas específicas y de error.*

Análisis factorial «Q» Forma grupos de encuestados o casos basados en su similitud a una serie de características (véase también la discusión sobre el análisis cluster en el Capítulo 9).

Análisis factorial «R» Analiza relaciones entre las variables para identificar grupos de variables que forman dimensiones latentes (*factores*).

Análisis *Jackknife* Procedimiento para deducir muestras repetidas basado en la omisión de una observación cuando se crea cada muestra. Permite la estimación empírica de los niveles de confianza del parámetro en lugar de la estimación paramétrica.

Análisis logit Véase regresión logística.

Análisis multivariante Análisis de varias variables en una única relación o un conjunto de relaciones.

Análisis univariante de la varianza (ANOVA) Técnica estadística que sirve para determinar, sobre la base de una medida dependiente, si las muestras provienen de poblaciones con igual media.

Aprendizaje Proceso secuencial de grandes ejemplos de observación (conocido como el ejemplo del *entrenamiento*) en el que se utilizan los errores de predicción o clasificación para volver a calcular los pesos para mejorar la estimación.

Árbol de decisión Modelo basado en reglas que consisten en *nodos* (puntos de decisión) y ramas (conexiones entre nodos) que alcanzan múltiples resultados al estar de paso entre dos o más nodos.

Ausencia aleatoria (MAR) Clasificación de los *datos ausentes* aplicable cuando los valores ausentes de Y dependan de X, pero no de Y. Cuando los datos ausentes son MAR, los valores observados para Y representan una muestra verdaderamente aleatoria para los valores X en la muestra pero no una muestra aleatoria de todos los valores Y debido a los valores ausentes de X.

Ausencia totalmente aleatoria (MCAR) Clasificación de los datos ausentes aplicable cuando los valores ausentes de Y no son dependientes de X. Cuando los datos ausentes son MCAR, los valores observados de Y son una muestra verdaderamente aleatoria de todos los valores de Y, sin procesos subyacentes que tiendan a sesgar los datos observados.

Autovalor También conocido como valor característico, una medida de la cantidad de varianza contenida en la matriz de correlación de tal forma que la suma de los autovalores sea igual al número de variables. También conocido como la raíz latente o raíz característica. Se calcula como la suma en columna de las cargas cuadradas para un factor. Representa la cantidad de varianza representada por un factor.

Autovalores Véase *raíces canónicas.*

Beta (β) Véase *error de Tipo II.*

Bootstrapping Forma de repetir la muestra en la que se extraen varias muestras de los datos originales con sustituciones para la estimación del modelo. Las estimaciones de los parámetros y los errores estándar no se calculan bajo supuestos estadísticos, sino a partir de observaciones empíricas.

Búsqueda concreta Utilizar datos más detallados usados para crear *datos agregados*. Por ejemplo, después de revisar las cifras de ventas semanales de un conjunto de tiendas, el investigador haría una búsqueda concreta para revisar los informes diarios de ventas o los informes de cada una de las tiendas. Se puede hacer una búsqueda al nivel más elemental de detalle (*datos originales*) de los elementos de los datos.

Calidad del ajuste Grado en el que la matriz de entradas —inputs— efectivas u observadas (covarianzas o correlaciones) se predice por el modelo estimado. Las medidas de *calidad del ajuste* se calculan sólo para la matriz de input total, no haciendo distinciones entre los constructos o indicadores exógenos o endógenos.

Capa oculta Series (capas) de *nodos* de una *red neuronal de perceptor multicapa* que están en los nodos input y output. El investigador puede o no controlar el número de nodos por capa oculta o el número de capas ocultas. Las capas ocultas ofrecen la capacidad de representar funciones no lineales en el sistema de redes neuronales.

Cargas canónicas También denominadas correlaciones de estructura canónicas, que miden la correlación lineal simple entre las variables independientes y sus respectivos valores teóricos canónicos y pueden ser interpretadas como cargas factoriales.

Cargas de los factores Una correlación entre las variables originales y los factores, y la clave para entender la naturaleza de un factor específico. Las cargas de los factores al cuadrado indican qué porcentaje de la varianza en una variable original se atribuye a un factor.

Cargas discriminantes Medida de la correlación lineal simple entre cada variable independiente y la puntuación Z discriminante para cada función discriminante, también denominadas correlaciones de estructura. Las cargas discriminantes se calculan tanto si se incluye la variable independiente en la función discriminante como si no.

Caso atípico En sentido estricto, una observación que tiene una diferencia sustancial entre el valor real de la variable criterio y el valor previsto. Casos que son sustancialmente «diferentes», tanto con relación a la variable criterio o independiente, se denominan habitualmente atípicos. El objetivo de detectar los casos atípicos es identificar las observaciones que no son representaciones apropiadas de la población de la cual se extrae la muestra, de forma que puedan ser descontados o incluso eliminados del análisis por no ser representativos.

Casos Heywood Un tipo habitual de *estimaciones infractoras*, que se producen cuando el término de error estimado para un indicador es negativo, que es un valor sin sentido. El problema se soluciona bien por la destrucción del indicador o bien por la restricción del valor de error de medida a un valor positivo pequeño.

Categoría de referencia El nivel omitido de una variable no métrica cuando se forma *una variable ficticia* de la variable no métrica.

Causación Principio por el cual se establecen las relaciones de causa efecto entre dos variables. Requiere que exista un grado suficiente de asociación (correlación) entre las dos variables, que una variable se produzca antes que la otra (que una variable sea claramente el resultado de la otra) y que no existan otras causas razonables para el resultado. Aunque en sus términos más estrictos raramente se encuentra la causación, en la práctica un fuerte soporte teórico puede hacer posible la estimación empírica de la causación.

Centroide Valor medio para la puntuación Z discriminante de una categoría particular o de un grupo. Por ejemplo, el análisis discriminante de dos grupos tiene dos centroides, uno para cada uno de los grupos. También puede ser la media o valor medio de los objetos contenidos en el conglomerado para cada variable, tanto si se utiliza en el *valor teórico cluster* o en el proceso de validación.

Codificación de efectos Método de especificación de la categoría de referencia para un conjunto de *variables ficticias* donde la categoría de referencia recibe un valor de menos uno (−1) a lo largo del conjunto de variables ficticias. Con este tipo de codificación, los coeficientes de las variables ficticias se convierten en desviaciones de grupo respecto a la media de todos los grupos. Se diferencia del *coeficiente de indicador*, donde a la referencia de categoría se le da el valor cero a lo largo de todas las variables ficticias y los coeficientes representan desviaciones de grupo respecto al grupo de referencia.

Codificación de indicadores El método para especificar la categoría de referencia para una serie de *variables ficticias* donde la categoría de referencia recibe un valor de cero a lo largo de la serie de variables ficticias. Los coeficientes de variables ficticias representan las diferencias de categoría respecto a la categoría de referencia. Véase también *codificación de efectos*.

Codificar Conversión de las categorías de variables no métricas en una serie de variables binarias, una para cada categoría. Parecido a la variable ficticia discutida en el Capítulo 2 sin la eliminación de la categoría de referencia.

Coeficiente Beta (B_{β}) Coeficiente de regresión tipificado (véase *tipificación*) que permite una comparación directa entre coeficientes en relación con su capacidad explicativa relativa de la variable criterio. Mientras los coeficientes de regresión se expresan en términos de unidades de la variable asociada, realizando por tanto comparaciones inapropiadas, los coeficientes beta utilizan datos tipificados y pueden ser comparados directamente.

Coeficiente de correlación (r) Indica la fuerza de la asociación entre las variables independientes y la variable criterio. El signo (+ o −) indica la dirección de la relación. Puede tomar valores entre −1 y +1, con +1 indicando una relación positiva perfecta, 0 indicando la ausencia de relación y −1 indicando una relación inversa o negativa perfecta (a medida que una crece, otra disminuye).

Coeficiente de correlación parcial Medidas de la fuerza de la relación entre el criterio o variable criterio y una variable predictor simple donde los efectos de la otra variable predictor del modelo se mantienen constantes. Por ejemplo, r_{Y1}, X_2, X_1, mide la variación en Y asociada con X_2 cuando el efecto de X_1 sobre X_2 e Y es constante. Usado en la estimación del modelo en el método de regresión para la selección secuencial de una variable, con el fin de identificar la variable independiente con el mayor poder predictivo *incrementado* por encima de las variables predictor ya presentes en el modelo.

Coeficiente de determinación (R^2) Medida de la proporción de la varianza de la variable criterio sobre su media, que es explicada por las variables independientes o predictor. El coeficiente puede variar entre 0 y 1. Si el modelo de regresión es estimado y aplicado apropiadamente, el investigador puede asumir que cuanto mayor

sea el valor de R^2, mayor será el poder explicativo de la ecuación de regresión, y por tanto mejor la predicción de la variable criterio.

Coeficiente de determinación corregido (R^2 corregido) Medida modificada del *coeficiente de determinación* que tiene en cuenta el número de variables predictor incluidas en la ecuación de regresión. Aunque la adición de variables a la regresión siempre provocará el aumento del coeficiente de determinación, el coeficiente de determinación corregido puede caer si las variables independientes añadidas tienen escaso poder explicativo y/o si los *grados de libertad* se vuelven demasiado pequeños. Este estadístico es muy útil para comparar entre ecuaciones con un número diferente de variables independientes, o tamaños muestrales diferentes, o ambos.

Coeficiente de regresión (b_n) Valor numérico del **parámetro** estimado directamente asociado con las variables independientes; por ejemplo, en el modelo $Y = b_0 + b_1X_1$, el valor b_1 es el coeficiente de regresión de la variable X_1. En el modelo de regresión múltiple (es decir, $Y = b_0 + +b_1X_1+b_2X_2$), los coeficientes de regresión son parciales porque cada uno tiene en cuenta no sólo las relaciones entre Y y X_1 y entre Y y X_2, sino también entre X_1 y X_2. El coeficiente no está limitado en su rango, ya que está basado tanto en el grado de asociación como en las unidades de escala de la variable predictor. Por ejemplo, dos variables con la misma asociación a Y tendrían diferentes coeficientes si una variable predictor se midió en una escala sobre 7 puntos y otra basada en una escala de 100 puntos.

Coeficiente discriminante Véase *ponderación discriminante*.

Coeficiente logístico Coeficiente que actúa como el factor de ponderación para las variables independientes en relación a su capacidad discriminatoria. Similar a una ponderación de regresión o coeficiente discriminante.

Colinealidad Expresión de la relación entre dos (colinealidad) o más (multicolinealidad) variables independientes. Se dice que dos variables predictor exhiben una completa colinealidad si su coeficiente de correlación es 1 y una completa falta de colinealidad si su coeficiente de correlación es 0. La *multicolinealidad* aparece cuando una única variable predictor está altamente correlacionada con un conjunto de otras variables independientes. Un caso extremo de la colinealidad/multicolinealidad es la *singularidad,* en la que se puede predecir una variable independiente perfectamente (es decir, correlación de 1,0) por otra variable independiente (de más de uno).

Combinaciones lineales Véase *valores teóricos canónicos*.

Comparación planificada Un *contraste a priori* que contrasta una comparación específica de las diferencias de las medias de los grupos. Estos contrastes son llevados a cabo conjuntamente con los contrastes para los *efectos interacción* y *principal* con la utilización de un *contraste*.

Componente parcial de la utilidad total Estimación a partir del análisis conjunto de la preferencia global o *utilidad* asociada con cada *nivel* de cada *factor* utilizado para definir el producto o servicio.

Comunalidad Varianza compartida con otras variables en el análisis factorial.

Condición de orden Requisito para la *identificación* según la cual los *grados de libertad* del modelo deben ser mayores o iguales que cero.

Condición de rango Requisito para la *identificación* en el que cada parámetro estimado está definido algebraicamente. Mucho más laboriosa que la *condición de orden* o estimación empírica de la identificación.

Confianza La proporción de veces que dos o más eventos se producen conjuntamente. Por ejemplo, si el producto A se compra el 60 por ciento de las veces que un cliente compra también el producto B, tenemos una confianza del 60 por ciento al decir que estos productos A y B se compran conjuntamente.

Conglomerado subjetivo Véase *datos confundidos*.

Conjunto de elección Conjunto de estímulos de perfiles completos construido a partir de los principios del diseño experimental y utilizado en el *enfoque basado en la elección*.

Constante (b_0) Valor del eje Y (eje de la variable criterio) donde la línea definida por la ecuación de regresión $Y = = b_0 + b_1X_1$ corta el eje. Se describe por el término constante b_0 en la ecuación de regresión. Además de su papel en la predicción, la constante puede o no tener una interpretación práctica. La constante representa el cambio en la variable criterio en ausencia de la variable predictor. Por ejemplo, cuando estimamos las ventas respecto de los gastos previos de publicidad, la constante representa el nivel de ventas esperado si se elimina el efecto de la publicidad. Pero en muchos casos, la constante sólo tiene valor predictivo dado que no existe situación en la cual están ausentes todas las variables predictor. Un ejemplo es la predicción de la preferencia por el producto basándose en las actitudes del consumidor. Todos los individuos tienen una actitud, de tal forma que la constante no tiene un uso para la toma de decisiones pero ayuda a realizar una predicción más precisa.

Constructo Concepto que el investigador puede definir en términos conceptuales pero que no puede ser directamente medido (es decir, el encuestado no puede articular una respuesta única que proporcionaría total y perfectamente una medida del concepto) o medido sin error (véase *error de medida*). Los constructos son las bases para formar *relaciones causales*, en la medida en que son las representaciones más «puras» posibles de un concepto. Un constructo puede definirse con diversos grados de precisión, que van desde los conceptos más simples, tales como la renta familiar, a los conceptos más complejos y abstractos, como la inteligencia y las emociones. Cualquiera que sea su nivel de especificidad, sin embargo, un constructo no puede ser medido directa y perfectamente aunque debe medirse aproximadamente por *indicadores*.

Constructo endógeno *Constructo* o variable que es la variable dependiente o resultante en al menos una *relación causal*. En términos de un diagrama de secuencias, existen una o más flechas apuntando *al* constructo o variable endógena.

Constructo exógeno *Constructo* o variable que actúa sólo como un predictor o «causa» para otros constructos o variables del modelo. En los diagramas de secuencias, los *constructos* exógenos sólo tienen flechas causales apuntando hacia fuera de ellos y no se pueden predecir por ninguno de los otros *constructos* del modelo.

Contraste Procedimiento para detectar diferencias entre los grupos específicos de interés junto con el ANOVA y el MANOVA —por ejemplo, comparando las diferencias de las medias de los grupos para un par de grupos específicos.

Contraste a priori Véase *comparación planificada.*

Contraste de esfericidad de Bartlett Prueba estadística de la significación general de todas las correlaciones con una matriz de correlación.

Contraste post hoc Contraste de las diferencias de medias realizado después de que los contrastes para los *efectos principales* se han realizado. Por lo general, los contrastes post hoc no utilizan un único *contraste,* sino que prueban las diferencias entre todas las combinaciones posibles de los grupos. Aunque proporcionan una información diagnóstica abundante, inflan la tasa de error de *Tipo I* con el empleo de contrastes estadísticos múltiples y por tanto tiene que utilizar niveles de confianza muy estrictos.

Contraste *t* Contraste para valorar la significación estadística de las diferencias entre dos medias muestrales para una única variable dependiente. Un caso especial del ANOVA para dos grupos o niveles de la variable de tratamiento.

Contrate M de Box Contraste estadístico para la igualdad de las matrices de varianzas/covarianzas de las variables dependientes entre los grupos. El contraste es muy sensible, especialmente a la presencia de una o más variables no-normales. Se emplea un nivel de significación de 0,01 o menor como ajuste para la sensibilidad del estadístico.

Correlación biserial Medida de correlación utilizada para reemplazar la correlación de momento-producto cuando una variable medida métricamente está asociada con una medida binaria (0,1) no métrica. Véase también *correlación poliserial.*

Correlación canónica Medida de la validez de la relación global entre las combinaciones lineales de los conjuntos de variables predictoras y criterio. Efectivamente, representa la correlación bivariante entre las dos combinaciones lineales.

Correlación del entorno Véase *correlación inter-atributos.*

Correlación inter-atributos Correlación entre atributos, también conocida como correlación de entorno que hace improbables o redundantes las combinaciones de atributos. Una correlación negativa representa esta situación en la cual dos atributos se supone que operan naturalmente en diferentes direcciones, tal como caballos de potencia y consumo de gasolina. Cuando uno aumenta, el otro se supone que naturalmente disminuye. Por tanto, a causa de esta correlación, todas las combinaciones de estos dos atributos (es decir, altos consumos de gasolina y gran potencia) no son creíbles. El mismo efecto se puede observar en correlaciones positivas, donde quizá se puede suponer que precio y calidad están positivamente correlacionados. No es creíble encontrar un producto de precio elevado y baja calidad en esta situación. La presencia de fuertes correlaciones inter-atributos requiere que el investigador examine detenidamente los estímulos presentados a los encuestados y evitar combinaciones increíbles que no son útiles en la estimación de los componentes parciales de la utilidad total.

Correlación parcial bivariante Correlaciones simples (dos variables) entre dos series de residuos (variación no explicada) que quedan tras eliminar la influencia de otras variables independientes.

Correlación policórica Medida de asociación empleada como sustituto de las correlaciones producto-momento cuando ambas variables sean medidas ordinales con tres o más categorías.

Correlación poliserial Medida de asociación utilizada como sustituto para la correlación producto-momento cuando se mide una variable en una escala ordinal y la otra variable se mide métricamente.

Correlación semiparcial Véase *coeficiente de correlación.*

Correlación tetracórica Medida de asociación utilizada para relacionar dos medidas binarias. Véase también *correlación policórica.*

Correlaciones de estructura Véase *cargas discriminantes.*

Covarianzas o análisis de covarianzas Empleo de procedimientos similares a la regresión para eliminar la variación extraña (molesta) de las variables dependientes debida a una o más variables independientes métricas incontroladas (covariaciones). Se supone que las covariaciones están linealmente relacionadas con las variables dependientes. Después del ajuste debido a la influencia de las covariaciones, se realiza un ANOVA (o MANOVA) estándar. Este proceso de ajuste (conocido como ANCOVA y MANCOVA) generalmente permite el uso de contrastes más sensibles de los efectos de tratamientos.

COVRATIO Medida de la influencia de una única observación sobre el conjunto completo de coeficientes de regresión estimados. Un valor cercano a 1 indica una escasa influencia. Si el valor de la COVRATIO menos 1 es mayor que $\pm 3p/n$ (donde p es el número de predictores y n es el tamaño muestral), la observación es considerada influyente basándonos en esta medida.

Criterio de aleatoriedad proporcional Otro criterio para calcular el ratio de aciertos, en donde se calcula la probabilidad «media» de la clasificación considerando los tamaños de todos los grupos.

Criterio de máxima aleatoriedad Medida de la exactitud predictiva de la matriz de clasificación que compara el porcentaje correctamente clasificado (también conocido como ratio de aciertos) con el porcentaje de encuestados en el grupo más grande. Su razón de ser está en que la mejor elección sin ningún tipo de información es clasificar cada observación dentro del grupo más grande.

Cuasi ortogonal Característica de un *diseño* de estímulos que no es *ortogonal,* pero que las desviaciones de la ortogonalidad son mínimas y están cuidadosamente controladas en la generación de los estímulos. Este tipo de diseño puede compararse con otros diseños de estímulos con medidas de *eficacia del diseño.*

Curtosis Medida del apuntamiento o llanura de una distribución cuando se compara con una distribución normal. Un valor positivo indica una distribución relativamente puntiaguda y un valor negativo indica una distribución relativamente plana.

Curva logística Una curva en forma de S obtenida por la transformación logit que representa la probabilidad de un suceso. La forma de S es no lineal porque la probabilidad de un suceso S se debe aproximar a 0 o a 1 pero nunca ser mayor. Así, aunque existe un componente lineal en el rango medio, según las probabilidades se aproximen a las cotas inferior y superior (0 y 1) deben «aplanarse» y ser asintóticas en estas cotas.

Chi-cuadrado Método de estandarización de datos en una *tabla de contingencia* comparando la frecuencia de la celda efectiva con la frecuencia de la celda esperada. La frecuencia de la celda esperada se basa en las probabilidades marginales de su fila y columna (probabilidad de una fila y columna entre todas las filas y columnas).

Datos agregados Datos creados a través de un proceso de combinación o sumatorio.

Datos analíticos Datos operacionales que han sido integrados y combinados con los datos analíticos existentes a través de la operación de *almacenamiento de datos* para formar un registro histórico de un objeto o evento de una empresa (es decir, cliente, empresa o producto).

Datos ausentes Información no disponible para un sujeto (o caso) sobre el cual existe otra información. Los datos ausentes a menudo tienen lugar cuando un encuestado falla en responder una o más cuestiones en una encuesta.

Datos ausentes prescindibles *Procesos de datos ausentes* que son explícitamente identificables o que están bajo el control del investigador. Los datos ausentes prescindibles no requieren una solución porque son inherentes a la técnica utilizada.

Datos censurados Observaciones que están incompletas de una manera sistemática y conocida. Por ejemplo, la investigación de las causas de la muerte en una muestra en la que algunos de los individuos todavía viven. Los datos censurados son un ejemplo de *datos ausentes prescindibles*.

Datos confundidos Procedimiento para obtener percepciones de encuestados de *datos de similitudes*. Los encuestados indican las similitudes entre pares de estímulos. La asociación (o «confusión») de unos estímulos con otros se considera que indica similitud. También son conocidos como *conglomerados subjetivos*.

Datos de similitudes Datos utilizados para determinar qué *objetos* son los más parecidos entre sí y cuáles son los más diferenciados. Implícita en las medidas de similitud está la capacidad de comparar todos los pares de objetos. Hay tres procedimientos para obtener datos de similitud, comparación por pares de objetos, *datos confundidos* y *medidas derivadas*.

Datos métricos También denominados *datos cuantitativos*, datos de *intervalos* o de *ratios*, estas medidas identifican o describen sujetos (u objetos) no sólo por la posesión de un atributo sino también por la cantidad o grado en que el sujeto está caracterizado por el atributo. Por ejemplo, la edad de una persona o su peso son medidas métricas.

Datos multidimensionales *Elementos de los datos* con múltiples *dimensiones* (atributos).

Datos no métricos También denominados *datos cualitativos*, son atributos, características o propiedades categóricas que identifican o describen a un sujeto o a un objeto.

Datos operacionales Datos utilizados por sistemas operacionales (es decir, contabilidad, gestión de existencias) que reflejan el estado en ese momento de la organización que es necesario para la gestión del día a día.

Datos originales Elementos de datos mantenidos a su nivel más bajo de detalle, tales como las transacciones individuales.

Definición conceptual Especificación de la base teórica para un concepto representado por un factor.

Dendrograma Representación gráfica (tres gráficos) de los resultados de un *procedimiento jerárquico* en el que cada *objeto* se representa en un eje, representando el otro eje los pasos del procedimiento jerárquico. Comenzando con cada objeto representado como un conglomerado separado, el dendrograma muestra gráficamente cómo se combinan los conglomerados en cada paso del procedimiento hasta que se encuentren todos contenidos en un conglomerado aislado.

Depósito Colección centralizada de *metadatos* disponibles para su acceso desde todas las funciones de *almacenamiento de datos*.

Descubrimiento de conocimiento en las bases de datos (KDD) Extracción de nueva información de las bases de datos a través de una variedad de procesos de descubrimiento de la información.

Desigualdad de Bonferroni Enfoque para ajustar el *nivel de significación alfa* seleccionado con el fin de controlar el porcentaje del *error de Tipo I* global. El procedimiento (1) incluye el cálculo del porcentaje ajustado *alfa* dividido por el número de contrastes que se van a realizar, y (2) emplea este porcentaje ajustado como valor crítico para cada uno de los contrastes.

DFBETA Medida de cambio en un coeficiente de regresión cuando se omite una observación del análisis de regresión. El valor de DFBETA es en términos del coeficiente mismo aunque también existe una versión estandarizada (SDFBETA). No pueden establecerse límites inferiores, aunque el investigador puede buscar valores sustancialmente diferentes del resto de las observaciones para evaluar la influencia potencial. Los valores de SDFBETA están determinados por sus errores estándares, por lo que apoya la racional para límites de 1 o 2, que corresponde a los niveles de confianza de 0,10 o 0,05 respectivamente.

DFFIT Medida del impacto de una observación en el ajuste conjunto del modelo o su versión estandarizada (SDFFIT). La mejor regla es identificar como influyente cualquier valor estandarizado (SDFFIT) que exceda $2/(\sqrt{p/n})$, donde p es el número de predictores y n es el tamaño muestral. No hay un valor límite para la medida DFFIT.

Diagrama de carámbanos en vertical Representación gráfica de conglomerados. Los *objetos* aislados se muestran horizontalmente en la parte de arriba del diagrama, y el proceso de agrupación jerárquico se representa en combinaciones de conglomerados verticalmente. Este diagra-

ma es parecido a un *dendrograma* invertido y ayuda en la determinación del número apropiado de conglomerados de la solución.

Diagrama de perfil Representación gráfica de los datos que ayuda en la identificación de los atípicos o la interpretación de la solución de conglomerado final. Normalmente, las variables del *valor teórico cluster* o aquellos utilizados para la validación se enumeran a lo largo del eje horizontal, y la escala de valor se utiliza para el eje vertical. Las líneas representan las puntuaciones (originales o estandarizadas) para *objetos* individuales o *centroides de los grupos* en un plano gráfico.

Diagrama de secuencias Representación gráfica de un conjunto completo de relaciones entre los constructos del modelo. Las *relaciones causales* se representan por flechas directas, saliendo la flecha de la variable predictor y apuntando a la variable o *constructo* dependiente. Las flechas torcidas representan correlaciones entre constructos o indicadores, pero no hay ninguna causación implicada.

Diagrama de tallo y hojas Una variante del *Histograma*, que proporciona una representación visual de la distribución de la variable, así como una enumeración de los valores efectivos de los datos.

Dimensión Atributo de un *elemento de datos*, tales como la edad o género de un cliente o fecha de un registro histórico.

Dimensiones Características de un *objeto*. Un objeto particular puede ser concebido como poseyendo tanto dimensiones *percibidas* o *subjetivas* (es decir, caro, frágil) como dimensiones *objetivas* (es decir, color, precio, características).

Dimensión objetiva Características físicas o tangibles de un *objeto* que tienen una base objetiva de comparación. Por ejemplo, un producto tiene tamaño, forma, color, peso, etc.

Dimensión percibida Una asignación subjetiva de características por parte del encuestado acerca de un objeto que representa sus características intangibles. A modo de ejemplo tendríamos «calidad», «caro» y «buena apariencia». Estas dimensiones percibidas son únicas para cada encuestado individual y pueden guardar una escasa relación con las *dimensiones objetivas*.

Dimensión subjetiva Véase *dimensión percibida*.

Dimensionalidad inicial Un punto de partida en la selección de las mejores configuraciones espaciales de los datos. Antes de comenzar un procedimiento de análisis dimensional, el investigador debe especificar cuántas *dimensiones* o caracteres están representados en los datos.

Diseño Conjunto específico de estímulos conjuntos creados para mostrar las propiedades estadísticas específicas de *ortogonalidad* y *equilibrio*.

Diseño de experimentos Plan de investigación en el cual el investigador manipula o controla directamente una o más variables predictoras (véase *tratamiento*) y valora sus efectos sobre las variables dependientes. Ya habitual en la ciencia física, está ganando popularidad en las ciencias económicas y sociales. Por ejemplo, a los encuestados se les muestra anuncios publicitarios distintos que varían sistemáticamente en una característica, tales como diferentes pretensiones (emocional frente a racional) o tipos.

Diseño de puente *Diseño* de estímulos para un gran número de *factores* (atributos) en los que los atributos se parten en varios grupos reducidos. Cada grupo de atributos tiene ciertos atributos que están contenidos en otros grupos, de tal forma que los resultados de cada grupo se pueden combinar, o construir puentes.

Diseño equilibrado *Diseño* de estímulos en los que en cada nivel dentro de un *factor* aparece un número igual de veces.

Diseño factorial Método de diseño de *estímulos* para evaluación mediante la generación de *todas* las posibles combinaciones de *niveles*. Por ejemplo, un análisis conjunto de tres factores con tres niveles por factor resultaría en 27 combinaciones que podrían actuar como estímulos.

Diseño factorial fraccionado Aproximación, como alternativa al *diseño factorial*, que utiliza sólo un subconjunto de posibles estímulos necesarios para estimar los resultados basándose en una regla de composición supuesta. Su tarea principal es reducir el número de las evaluaciones recogidas mientras se mantiene la *ortogonalidad* entre los *niveles* y subsecuentes estimaciones de los *componentes parciales de la utilidad total*. El diseño más simple es un *modelo aditivo*, en el cual sólo se estiman los *efectos principales*. Si los *términos de interacción* están incluidos, entonces se crean estímulos adicionales. El diseño puede ser creado tanto por la relación con las fuentes publicadas como utilizando los programas informáticos que acompañan a la mayoría de los paquetes de análisis conjunto.

Diseño óptimo ortogonal Restricción matemática que se exige para que las estimaciones de los *componentes parciales de la utilidad total* sean independientes unas de las otras. En el análisis conjunto, la ortogonalidad se refiere a la capacidad para medir el efecto de cambiar cada nivel de atributo y separarlo de los efectos del cambio sobre otros niveles de atributos y de un error experimental.

Disparidades Diferencias en las distancias generadas por ordenador que representan similitudes y las distancias provistas por el encuestado.

Distancia de Cook (D_i) Medida resumen de la influencia de un único caso (observación) basado en el cambio total del resto de los residuos cuando se elimina el caso del proceso de estimación. Grandes valores (generalmente mayores que 1) indican una influencia sustancial del caso que afecta a los coeficientes de regresión estimados.

Distancia de Mahalanobis (D^2) Medida del impacto de un caso aislado basándose en la diferencia entre el valor del caso y el valor de la media del resto de los casos para todas las variables independientes. La fuente de influencia sobre los resultados de la regresión es para este caso muy diferente de la que ejercen una o varias variables predictor, por tanto provocando un desplazamiento de la ecuación de regresión. Es una forma estandarizada de la *distancia euclídea*, es decir, una medición de respuestas en términos de desviaciones estándar que estandarizan los datos, con el ajuste hecho para las intercorrelaciones entre las variables.

Distancia euclidea Medida de *similitud* entre dos *objetos* más utilizada. Esencialmente, es una medida de la longitud de una línea recta dibujada entre dos objetos.

Distribución normal Una distribución de probabilidad continua puramente teórica en la que el eje horizontal representa todos los posibles valores de una variable y el eje vertical representa la probabilidad de que ocurran esos valores. Los valores de la variable están agrupados en torno a la media de forma simétrica y unimodal, conocida como la curva en forma de campana de Gauss o normal.

Distribución normal multivariante Generalización de la distribución normal univariante para el caso de p variables. Una distribución normal multivariante de grupos muestrales es un supuesto básico requerido para la validez de los contrastes de significación en muchas técnicas multivariantes.

Efecto interacción En los diseños factoriales, los efectos conjuntos de dos variables de tratamiento junto a los *efectos principales* individuales. Esto significa que las diferencias entre grupos en una variable de tratamiento varían dependiendo del nivel en la segunda variable de tratamiento. Por ejemplo, supongamos que los encuestados son clasificados por la renta (tres niveles) y sexo (masculino y femenino). Se daría una interacción significativa cuando las diferencias en la(s) variable(s) independiente(s) entre masculino y femenino variase de forma sustancial entre los tres niveles de renta.

Efecto moderador Impacto de una tercera variable independiente (la variable moderador) que causa un cambio en la relación entre un par de variables dependiente/independiente en función del valor de la variable moderador. También conocida como efecto interactivo y similar al efecto interacción visto en el análisis de los métodos de la varianza.

Efecto principal En los diseños factoriales, el efecto individual de cada variable de *tratamiento* sobre la variable dependiente.

Efecto tamaño Medida estandarizada de las diferencias de los grupos empleada en el cálculo de la potencia estadística. Se calcula como la diferencia entre las medias de grupo dividido por la desviación estándar, entonces es comparable por todos los estudios de investigación como una medida de efecto generalizada (es decir, las diferencias entre las medias de grupo). Es una estimación del grado en que el fenómeno estudiado (por ejemplo, la correlación o la diferencia de medias) existe en la población.

Efecto tipo de respuesta Series de respuestas sistemáticas de un encuestado que reflejan un «sesgo» o pauta consistente. Los ejemplos más claros con las respuestas sobre un objeto que siempre son excelentes o pésimas para todos los atributos con escasa o nula variación.

Eficacia del diseño Grado en que el *diseño* se ajusta al diseño *ortogonal*. Esta medida se utiliza principalmente para evaluar y comparar diseños *cuasi-ortogonales*. Los valores de la eficacia del diseño van de 0 a 100, que denota un diseño *óptimo*.

Elementos de los datos Nivel de datos almacenados en una base de datos. Los elementos de datos aislados pueden representar diferentes niveles de agregación o suma, que van desde los *datos originales* a los *datos agregados*.

Eliminación de datos Véase *limpieza de datos*.

Eliminación progresiva (*Backward*) Método de selección de variables para incluirlas en el análisis de regresión que comienza con todas las variables independientes en el modelo y posteriormente va eliminando aquellas variables que no suponen una contribución significativa a la predicción.

Encadenamiento completo *Algoritmo de aglomeración* en el que la *similitud entre objetos* se basa en la distancia máxima entre *objetos* en dos conglomerados (la distancia entre los miembros más diferentes de cada conglomerado). A cada paso de la aglomeración, se combinan los dos conglomerados con la distancia máxima más pequeña (más parecidos).

Encadenamiento medio *Algoritmo* utilizado en los *métodos de aglomeración* que representa la *similitud* como la distancia media de todos los objetos dentro de un conglomerado a todos los objetos de otro. Esta aproximación tiende a combinar los conglomerados con varianzas reducidas.

Encadenamiento simple Procedimiento de *análisis cluster jerárquico* en el cual se define la *similitud* como la distancia mínima entre cualquier *objeto* de un conglomerado y cualquier objeto de otro. Esto simplemente es la distancia entre los objetos más cercanos de dos conglomerados. Este procedimiento tiene el potencial de creación de conglomerados menos compactos e incluso encadenados. Difiere del método de *encadenamiento completo*, que utiliza la distancia máxima entre los objetos del conglomerado.

Enfoque conjunto basado en la elección Forma alternativa de recogida de respuestas y estimación del modelo conjunto. La diferencia principal está en que los encuestados seleccionan un único *estímulo de perfiles completos* de un conjunto de estímulos (conocido como *conjunto de elección*) en lugar de clasificar o calificar los estímulos por separado.

Enfoque de disponibilidad completa *La técnica de imputación* para datos ausentes que calcula los valores basándose en todas las observaciones válidas disponibles.

Enfoque de la distancia absoluta Método de calcular distancias basado en la suma de las diferencias absolutas de las coordenadas para los objetos. Este método supone que las variables están correlacionadas y que las unidades de medida son compatibles.

Enfoque de observaciones completas Técnica para tratar los datos ausentes que calcula valores basándose sólo en datos de casos completos, es decir, casos sin *datos ausentes*.

Error de especificación Error de predicción de la variable criterio provocado por la exclusión de una o más variables independientes relevantes que pueden sesgar los coeficientes estimados de las variables incluidas así como una disminución del poder predictivo conjunto del modelo de regresión. Señala la falta de calidad del ajuste de un modelo debido a la omisión de una variable relevante del modelo propuesto. Los contrastes para los errores de es-

pecificación son bastante complicados y requieren numerosos intentos con modelos alternativos. El investigador puede evitar el error de especificación en gran medida utilizando sólo bases teóricas para la construcción del modelo propuesto. De esta forma, el investigador está menos expuesto a «dejar pasar» un constructo relevante para el modelo.

Error de medida Grado en que los valores de los datos no están midiendo verdaderamente la característica de estar representados por la variable. Por ejemplo, cuando preguntamos sobre la renta total familiar, existen muchas fuentes de medida del error (resistencia a contestar la cantidad completa, error en la estimación de la renta total, etc.,) que hacen imprecisos los valores de los datos. Representa el grado en el que las variables que podemos medir (las *variables manifiestas*) no describen perfectamente lo(s) *constructo(s) latente(s)* de interés. Las fuentes de error de medida pueden ir de simples errores de entrada de datos a la definición de constructos (por ejemplo, conceptos abstractos tales como patriotismo o lealtad que quieren decir cosas distintas para diferentes personas) que no están perfectamente definidos por ningún conjunto de variables manifiestas. A efectos prácticos, todos los constructos tienen cierto error de medida, incluso con las mejores *variables indicador*. Sin embargo, el objetivo del investigador es minimizar la cantidad del error medido. SEM puede tomar en cuenta el error de medida a efectos de proporcionar estimaciones más precisas de las *relaciones causales*.

Error de predicción Diferencia entre los valores reales y de predicción de la variable criterio para cada observación en la muestra (véase *residuo).*

Error de Tipo I Probabilidad de rechazar la hipótesis nula siendo cierta, es decir, concluir que dos medias son significativamente diferentes cuando de hecho son iguales. Valores pequeños de alfa (por ejemplo, 0,05 ó 0,01), (véase *alfa*), llevan a que el rechazo de la hipótesis nula sea insostenible y a la aceptación de la hipótesis alternativa de que las medias poblacionales son distintas.

Error de Tipo II Probabilidad de no rechazar la hipótesis nula cuando es falsa rechazada, es decir, concluir que las dos medias no son significativamente diferentes cuando de hecho lo son. También conocido como error *beta* (β). En términos más sencillos, la posibilidad de no encontrar correlación o diferencia de medias cuando existen. Está inversamente relacionado con el *error de Tipo I.* El valor 1 menos el error de Tipo II se define como la *potencia.*

Error estándar Medida de la dispersión de las medias o de las diferencias de las medias esperada debido a la variación muestral. Denota la gama esperada del coeficiente a traves de las muestras múltiples de los datos. Resulta de utilidad en pruebas estadísticas de significación para ver si el coeficiente es sustancialmente diferente de cero (es decir, si la gama esperada del coeficiente contiene el valor cero a un nivel de confianza específico). El valor *t* de un *coeficiente de regresión* es el coeficiente dividido por su error estándar.

Error estándar de la estimación (SEE) Medida de la variación en el valor previsto que puede usarse para desarrollar intervalos de confianza alrededor de cualquier variable a predecir. Parecido a la desviación estándar de una variable alrededor de la media.

Escala de similitud Escala arbitraria, por ejemplo, desde −5 a +5 que permite la representación de una relación ordenada entre objetos desde los más similares (más parecidos) a los menos similares (menos parecidos). Este tipo de escala es apropiado sólo para representar una dimensión única.

Escalas aditivas Método de combinar varias variables que miden el mismo concepto en una única variable con el fin de aumentar la *fiabilidad* de la medida. En la mayoría de los casos, se suman las variables separadas y a continuación se usa su puntuación total o media en el análisis. También se conocen como escalas sumadas.

Espacio de búsqueda Rango de soluciones potenciales definidas por diversos modelos de aprendizaje, tales como *algoritmos genéticos* o *redes neuronales*. La solución se puede especificar en términos de las dimensiones consideradas, incluidos los elementos de datos o criterios para soluciones factibles o no factibles.

Estadístico de Wald Contraste utilizado en la regresión logística para la significación del coeficiente logístico. Su interpretación es como la de los valores *F* o *t* utilizados en los contrastes de significación de los coeficientes de regresión.

Estadístico PRESS Medida de validación obtenida eliminando una observación cada vez y predecir este valor dependiente con el modelo de regresión estimado para el resto de las observaciones.

Estadístico *Q* de press Medida de la capacidad clasificatoria de la función discriminante cuando se compara con los resultados esperados de un modelo aleatorio. El valor calculado se compara con un valor crítico basado en una distribución chi-cuadrado. Si el valor calculado excede al valor crítico, los resultados de la clasificación son significativamente mejores que lo que se esperaría de forma aleatoria.

Estadístico *t* Contraste estadístico que valora la significación estadística entre dos grupos en una única variable dependiente (véase *contraste t*).

Estadístico *U* Véase *lambda de Wilks.*

Estimación de máxima verosimilitud (MLE) Método de estimación habitualmente empleado en modelos de ecuaciones estructurales, incluyendo LISREL y EQS. Método alternativo a los mínimos cuadrados ordinarios utilizados en la regresión múltiple, MLE es un procedimiento que mejora iterativamente los parámetros estimados para minimizar una función de ajuste especificada.

Estimación directa Proceso por el cual un modelo se estima directamente con un procedimiento de estimación seleccionado y el intervalo de confianza (y el error estándar) de cada parámetro estimado se basa en el error muestral.

Estimación por etapas Proceso de estimación de la(s) función(es) discriminante(s) o del modelo de regresión logística según el cual las variables independientes se introducen secuencialmente de acuerdo a la capacidad discriminatoria que añaden a la predicción de pertenencia al grupo.

Estimación por pasos en la regresión múltiple Método de selección de variables para su inclusión en el modelo de regresión que comienza con la selección del mejor predictor de la variable criterio. Las variables independientes adicionales se seleccionan en términos del aumento del poder explicativo que puede añadirse al modelo de regresión. Las variables independientes se añaden a medida que sus *coeficientes de correlación parcial* son estadísticamente significativos. Se pueden dejar de usar las variables independientes si su poder predictivo cae a un nivel de no significatividad cuando se añade otra variable independiente al modelo.

Estimaciones infractoras Cualquier valor que exceda sus límites teóricos. Las más habituales son los casos *Heywood* con errores de varianza negativos (el valor mínimo debería ser cero, indicando que no existe un error de medida) o errores estándar muy grandes. El investigador debe corregir la estimación infractora con uno de los numerosos remedios antes de que los resultados puedan ser interpretados por el ajuste conjunto del modelo y los coeficientes individuales puedan ser examinados por su significación estadística.

Estímulo Conjunto específico de *niveles* (uno por *factor*) evaluados por los encuestados (también conocido como *tratamiento*). Se obtiene un método de definición de estímulos (*diseño factorial*) al tomar todas las combinaciones de todos los niveles. Por ejemplo, tres factores con dos niveles cada uno crearían 8 estímulos. Sin embargo, en muchos análisis conjuntos, el número total de combinaciones es muy grande para que un encuestado los evalúe todos. En estos casos, algunos conjuntos de estímulos se crean de acuerdo a un plan sistemático, generalmente un *diseño factorial fraccional.*

Estímulo de validación Conjunto de *estímulos* que no utilizan la estimación de los componentes parciales de la utilidad total. Estos componentes se utilizan para predecir preferencia en la validación de estímulos para asegurar la validez y la fiabilidad de las estimaciones originales. Parecido en su concepción a la muestra de validación de los encuestados en el análisis discriminante.

Estímulos de ampliación de muestra Véanse *estímulos de validación.*

Estrategia de desarrollo del modelo Estrategia de modelización estructural que incorpora *reespecificación de modelo* como un método teóricamente conducente a la mejora de un modelo especificado de forma tentativa. Esto permite la exploración de formulaciones alternativas del modelo que puedan ser apoyadas por la teoría. No corresponde a un *enfoque exploratorio* en el que las reespecificaciones del modelo se hacen ateóricamente.

Estrategia de modelos confirmatorios Estrategia que evalúa estadísticamente un único modelo por su ajuste a los datos observados. Este enfoque es en realidad menos riguroso que la *estrategia de modelos rivales* porque no considera modelos alternativos que podrían ajustarse mejor o igualmente bien al modelo propuesto.

Estrategia de modelos rivales Estrategia que compara el modelo propuesto con un número de modelos alternativos en un intento de demostrar que no existe un modelo mejor ajustado. Esto es particularmente relevante en un modelo de ecuaciones estructurales dado que un modelo puede mostrarse teniendo sólo un ajuste aceptable, pero sólo un ajuste aceptable no garantiza que otro modelo no se ajustará mejor o igualmente bien.

Estructura de preferencias Representación tanto de la importancia relativa o valor de cada *factor* y el impacto de los *niveles* individuales que afectan a la utilidad.

Evaluación subjetiva Método de determinación de cuántas *dimensiones* están representadas en el modelo de análisis multidimensional. El investigador hace una «inspección subjetiva» de los mapas espaciales y se pregunta si la configuración parece razonable. El objetivo es obtener el mejor ajuste con el menor número de dimensiones.

Exposición de preferencias Representación de las *preferencias* de un encuestado dentro de un espacio de estímulos comunes (agregado) derivado para todos los encuestados en conjunto. Las preferencias de los individuos están «expuestas» y retratadas como la mejor representación posible dentro del análisis agregado.

Extensión de los vectores El vector escalado en el que el vector original es escalado para representar el correspondiente ratio *F*. Se emplea para representar gráficamente la carga de la función discriminante de forma combinada con las centroides de grupo.

Extracción de datos Extraer información práctica válida y previamente desconocida de grandes bases de datos y aplicarla a modelos de empresa.

Factor (Análisis conjunto/MANOVA) Variable independiente no-métrica, también denominada variable de experimento o de *tratamiento*.

Factor (Análisis factorial) Combinación lineal (valor teórico) de las variables originales. Los factores también representan las dimensiones subyacentes (construcciones) que resumen o justifican la serie original de variables en observación.

Factor de inflación de varianza (VIF_i) Medida del efecto del resto de las variables predictor sobre un coeficiente de regresión. Directamente relacionada con el valor de *tolerancia* ($VIF_i = 1/TOL_i$). Altos valores VIF (con un límite habitual en 10,0) indican también un grado de *colinealidad* o *multicolinealidad* entre las variables independientes.

Factor en bloques Característica de los encuestados en el ANOVA/MANOVA que se utiliza para reducir la variabilidad dentro del grupo. Esta característica llega a ser un *tratamiento* adicional del análisis, con lo que se obtienen grupos adicionales que son más homogéneos. Como ejemplo, supongamos que se pregunta a los clientes su intención de compra de un producto y la medida independiente empleada es la edad. El examen de los datos demuestra que la variación principal se debe al sexo. Entonces el sexo podía ser añadido como un tratamiento adicional, por lo que cada categoría de edad sería dividida en los grupos masculino y femenino con mayor homogeneidad dentro de cada grupo.

Fiabilidad Grado en que una variable o conjunto de ellas es coherente con lo que pretende medirse. Si se toman múlti-

ples medidas, las medidas fiables serán muy coherentes en sus valores. Se diferencia de la *validez* en que no se relaciona con lo que debería medirse, sino con cómo se mide.

Fiabilidad (*Modelos de estructuras de covarianzas*) Grado en el que un conjunto de *indicadores de constructo latente* son consistentes en sus medidas. En términos más formales, la fiabilidad es el grado en que un conjunto de dos o más indicadores «coinciden» en sus medidas de un constructo. Los indicadores de constructos altamente fiables están altamente correlacionados, indicando que todos ellos están midiendo el mismo constructo latente. A medida que la fiabilidad disminuye, los indicadores se hacen menos consistentes y por tanto son peores indicadores del constructo latente. La fiabilidad puede calcularse como 1,0 menos el *error de medida*. Véase también *validez*.

Función canónica Relación (de correlación) entre dos combinaciones lineales. Cada función canónica tiene dos combinaciones lineales distintas (*valores teóricos canónicos*), una para el conjunto de variables criterio y otra para el conjunto de variables predictoras. La validez de la relación está dada por la *correlación canónica*.

Función de activación Función matemática del *nodo* que traslada la puntuación sumada de los valores de los *input* ponderados en un único valor *output*. Aunque la función puede ser en principio de cualquier tipo, la forma más común es una *función sigmoide*.

Función de base radial Forma alternativa de *red neuronal* que funciona más como un *perceptor multicapa*, siendo apropiada para los problemas de predicción y clasificación.

Función de clasificación Método de clasificación en el que se define una función lineal para cada grupo. La clasificación se lleva a cabo calculando una puntuación para cada observación sobre cada función de clasificación de cada grupo y asignando la observación al grupo con mayor puntuación. Esto difiere del cálculo de la *puntuación Z discriminante*, que se calcula por medio de la *función discriminante*.

Función de distancia normalizada Proceso que convierte cada puntuación de los datos iniciales en un valor teórico estandarizado con una media de 0 y una desviación estándar de 1, que elimina el sesgo introducido por las diferencias en las escalas de varias variables.

Función discriminante Un valor teórico de las variables independientes seleccionadas por su capacidad discriminante utilizadas en la predicción de la pertenencia al grupo. El valor predicho de la función discriminante es la puntuación *Z* discriminante que se calcula para cada objeto (persona, empresa o producto) en el análisis. Adopta la forma de la ecuación lineal:

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \cdots + W_n X_{nk}$$

donde

Z_{jk} = Puntuación *Z* discriminante de la función discriminante *j* para el objeto *k*.

a = Constante.

W_i = Ponderación discriminante para la variable independiente *i*.

X_{ik} = Variable independiente *i* para el objeto *k*.

Función discriminante lineal de Fisher Véase *función de clasificación*.

Función sigmoide Función no lineal con una distribución general en forma de S. Un ejemplo habitual es la función logística.

Grados de libertad El número de unidades de información disponibles para estimar la distribución muestral de los datos después de que se hayan estimado todos los parámetros. Calculado desde el número total de observaciones menos el número de *parámetros* estimados. En términos prácticos, los grados de libertad son el número de correlaciones no redundantes o covarianzas en la matriz de inputs menos el número de coeficientes estimados. Los investigadores intentan maximizar los grados de libertad disponibles, mientras obtengan el modelo mejor ajustado. Cada coeficiente estimado «gasta» un grado de libertad. Un modelo nunca puede estimar más coeficientes que el número de correlaciones no redundantes o covarianzas, lo que significa que cero es el nivel más bajo de los niveles de libertad para cualquier modelo. Estos parámetros estimados son restricciones de los datos, porque una vez estimados, definen la población de la cual se supone han sido extraídos los datos. Por ejemplo, al estimar un modelo de regresión múltiple con una única variable predictor, estimamos dos parámetros, la constante (b_0) y un coeficiente de regresión para la variable predictor (b_1). Al estimar el error aleatorio, definido como la suma de los errores de predicción (variables dependientes reales menos las predicciones) para todos los casos, encontraríamos (n – 2) grados de libertad. Los grados de libertad proporcionan una medida de la restricción de los datos a la hora de alcanzar un nivel de predicción. Si los grados de libertad son pequeños, esto sugiere que la predicción resultante puede ser menos generalizable, dado que todas las observaciones, con la excepción de unas pocas han sido incorporado a la predicción. Inversamente, un valor mayor de grados de libertad indica que la predicción es claramente "robusta" con relación a su representatividad del total de la muestra de encuestados.

Gráfico de cajas y bigotes Método de representación de la distribución de la variable. Una caja representa la mayor parte de la distribución y las extensiones, llamadas bigotes, llegan al final de la distribución. Muy útil para realizar comparaciones de una sola variable a lo largo de grupos o entre varias variables.

Gráfico de distribución normal Comparación gráfica de la forma de la distribución respecto a la *distribución normal*. En el gráfico, la distribución normal se representa por una línea recta con un ángulo de 45 grados. La distribución efectiva se dibuja contra esta línea, de tal forma que cualquier diferencia se muestra como desviación de la línea recta, haciendo muy aparente e interpretable la identificación de estas diferencias.

Gráfico de no-correlación de residuos Gráfico de residuos respecto al valor previsto que exhibe un patrón aleatorio. Un gráfico de no-correlación de residuos es un indicio de violaciones no identificables de los supuestos subyacentes en el análisis de regresión.

Gráfico de probabilidad normal Gráfico de comparación de la forma de la distribución de la muestra con la distribución normal. En el gráfico, la distribución normal se presenta por una línea recta en ángulo de 45 grados. La distribución efectiva se representa gráficamente respecto a esta línea, de tal forma que cualquier diferencia se muestra como una desviación de la línea recta, realizando así la identificación de las diferencias bastante simples.

Gráfico de regresión parcial Representación gráfica de la relación entre la variable criterio y una única variable independiente. El gráfico de dispersión de los puntos representa la correlación parcial entre las dos variables, con los efectos de otras variables independientes constantes (véase *coeficiente de correlación parcial*). Es particularmente útil en la evaluación de la forma de la relación (lineal o no lineal) y la identificación de *observaciones influyentes*.

Gráficos de dispersión Representación de las relaciones entre dos variables métricas que muestren los valores conjuntos de cada observación en un gráfico de dos dimensiones.

Grupo de comparación La categoría de una variable no métrica que recibe todos ceros cuando se usa la *codificación de indicadores* o todos menos unos (–1) cuando se usa la *codificación de efectos* para crear *variables ficticias*.

Grupo de entropía Grupo de *objetos* independiente de cualquier conglomerado (es decir, no se ajustan a ningún conglomerado) que pueden ser considerados como atípicos y posiblemente eliminados del análisis *cluster*.

Heterocedasticidad Véase *homocedasticidad*.

Hipótesis nula Hipótesis de que las muestras provienen de poblaciones con medias iguales para una variable dependiente (contraste univariante) o para un conjunto de variables dependientes (contraste multivariante). La hipótesis nula puede ser aceptada o rechazada dependiendo de los resultados del contraste de significación estadística.

Histograma Representación gráfica de la distribución de una única variable. Al formar el recuento de frecuencias en categorías, puede verse el perfil de la distribución de la variable. Utilizado para realiza una comparación visual con la *distribución normal*.

Homocedasticidad Descripción de datos en los que la varianza del término de error (ϵ) aparece constante sobre un rango de variables independiente. El supuesto de igual varianza del error de la población E (donde E se estima de ϵ) es decisivo para la apropiada aplicación de la regresión lineal. Cuando el término de error tiene una varianza en aumento u ondulante, se dice que los datos son *heterocedásticos*. La discusión de los *residuos* ilustra mejor este punto.

Identificación Grado en el que existe un número suficiente de ecuaciones a «resolver para» cada uno de los coeficientes (incógnitas) a estimar. Los modelos pueden estar *infraidentificados* (no se pueden resolver), *identificados* (número de ecuaciones que igual al número de coeficientes estimados sin grados de libertad) o *sobreidentificados* (más ecuaciones que coeficientes estimados y grados de libertad mayores que cero). El investigador desea tener un modelo sobreidentificado por el mayor rigor de los contrastes del modelo propuesto. Véase también *grados de libertad*.

Independencia Supuesto básico del ANOVA/MANOVA que requiere que las medidas dependientes para cada encuestado estén totalmente incorrelacionadas con las respuestas de otros encuestados de la muestra. La ausencia de independencia afecta de forma muy importante a la validez estadística de los análisis a menos que se lleve a cabo una corrección.

Indeterminación de factores Característica del *análisis factorial común* por el que se pueden calcular varias puntuaciones diferentes dadas por un encuestado, cada una adecuada al modelo factorial estimado. Esto significa que las puntuaciones de factores no son únicas para cada individuo.

Indicador Valor observado (*variable manifiesta*) utilizado como una medida de un concepto o *constructo latente* que no puede ser medido directamente. El investigador debe especificar que indicadores están asociados con cada *constructo latente*.

Índice de ajuste Índice de correlación al cuadrado (R^2) que puede ser interpretado como indicativo de la proporción de variación de las *disparidades* (datos óptimamente analizados) que pueden ser tenida en cuenta por el procedimiento del análisis multidimensional. Mide la bondad del ajuste de los datos al modelo multidimensional. Este índice es una alternativa a la *medida de stress* en la determinación del número de dimensiones. Similar a las medidas de la covarianza en otras técnicas multivariantes, medidas de 0,60 o superiores se consideran aceptables.

Índice de condicionamiento Medida de la proporción relativa de la varianza asociada con un *autovector,* de tal forma que un elevado índice de condicionamiento indica un elevado grado de *colinealidad.*

Índice de potencia Medida compuesta de la capacidad discriminatoria de una variable independiente cuando se estima más de una función discriminante. Basado en las cargas discriminantes, es una medida relativa con la que se pueden comparar la discriminación global proporcionada por cada una de las variables independientes a lo largo de todas las funciones discriminantes significativas.

Índice de redundancia Cantidad de varianza de un *valor teórico canónico* (dependiente o independiente) explicada por el otro *valor teórico canónico* de la función canónica. Se puede calcular tanto para valores teóricos canónicos dependientes como independientes en cada función canónica. Por ejemplo, un índice de redundancia de la variable dependiente representa la cantidad de varianza de la variable dependiente explicada por el valor teórico canónico independiente.

Índices de modificación Valores calculados para cada posible relación no estimada en un modelo especificado. El valor del índice de modificación para una relación concreta no estimada indica la mejora en un modelo de ajuste conjunto (la reducción en el estadístico chi-cuadrado) que es posible si un coeficiente se calcula para esa relación no contrastada. El investigador debería utilizar

los índices de modificación sólo como guía para mejoras del modelo en aquellas relaciones que puedan ser justificadas teóricamente como posibles modificaciones.

Inteligencia artificial (AI) Área general de la ciencia informática que trata de la creación de programas de ordenadores que intentan emular las propiedades de aprendizaje del cerebro humano. Posteriormente, la AI ha evolucionado hacia diversas áreas más especializadas, incluyendo las *redes neuronales* y los *algoritmos genéticos*.

Interacción disordinal Forma del *efecto interacción* entre las variables independientes que invalida la interpretación de los *efectos principales* de los tratamientos. Una interacción disordinal se muestra de forma gráfica representando las medias para cada grupo y considerando las líneas que se intersectan y cruzan. En este tipo de interacción no sólo varían las diferencias de las medias dadas las combinaciones únicas de los niveles de las variables independientes, sino que la ordenación relativa de los grupos también cambia.

Interacción ordinal Clase aceptable de un *efecto interacción* en el que las magnitudes de las diferencias entre los grupos varían, pero las posiciones relativas de los grupos permanecen constantes. Se representa gráficamente dibujando los valores medios y observando las líneas que no son paralelas, que no se intersectan.

Jitter Cantidades pequeñas de ruido blanco añadidas a los valores utilizados en el *entrenamiento* de una red neuronal, que suavizan la función de error y asisten en la obtención de una solución global óptima.

Lambda de Wilks Uno de los cuatro estadísticos fundamentales para contrastar la hipótesis nula en el MANOVA. También relacionado con el criterio de máxima verosimilitud o el *estadístico U*.

Lenguaje de petición estructurada (SQL) Método de extracción de información desde una base de datos de acuerdo a los criterios especificados. Difiere de la extracción de datos en que sólo los resultados ofrecidos son aquellos que cumplen las condiciones específicas requeridas. SQL no puede desarrollar ningún cálculo analítico, como desarrollar modelos de clasificación o determinar niveles de asociación. Por ejemplo, SQL puede extraer información de todos los clientes entre los 35 y los 55 años que compraron un producto en el último mes mediante una tarjeta de crédito, pero no puede agrupar a los clientes en segmentos sin criterios específicos.

Limpieza de datos Eliminación de los errores aleatorios y sistemáticos de los *elementos de datos* a través del filtrado, fusión y traslación.

Linealidad Utilizada para expresar el concepto de que el modelo posee las propiedades de aditividad y homogeneidad. En un sentido simple, los modelos lineales predicen valores que caen en una línea recta que tiene un cambio unitario constante de la variable dependiente (pendiente) a partir de un cambio unitario constante de la variable independiente. En el modelo de población $Y = b_0 + b_1X_1 + e$, el efecto de un cambio de 1 en X_1 es añadir b_1 (una constante) unidades de Y.

M de Box Contraste estadístico para la igualdad de las matrices de covarianzas de las variables independientes entre los grupos de la variable dependiente. Si la significación estadística es mayor que el nivel crítico (por ejemplo, el 0,01), esto apoya la igualdad de las matrices de covarianzas. Si el contraste muestra significación estadística, los grupos se juzgan diferentes y no se verifica el supuesto.

Mapa espacial Véase *mapa perceptual*.

Mapa perceptual Representación visual de las percepciones de un encuestado de *objetos* en dos o más dimensiones. Habitualmente este mapa tiene niveles opuestos de *dimensiones* de los ejes X e Y, tales como «dulce» y «amargo» al final de los ejes X y «caro» y «barato» al final del eje Y.

Mapa territorial Representación gráfica de las puntuaciones de corte sobre un gráfico bidimensional. Cuando se combinan con los puntos de los casos individuales, se puede ver la dispersión de cada grupo e identificarse directamente sobre el mapa la clasificación errónea de los casos individuales.

Matriz de clasificación Matriz que valora la capacidad predictiva de la(s) función(es) discriminante(s) o de la regresión logística. También denominada matriz de confusión, de asignación o de predicción, es una matriz que contiene los números que reflejan la capacidad predictiva de la función discriminante. Se crea tabulando de forma cruzada el miembro del grupo correcto con el miembro del grupo predicho, los números de la diagonal de la matriz representan clasificaciones correctas y los números fuera de la diagonal son clasificaciones incorrectas.

Matriz de correlación anti-imagen Una matriz de las correlaciones parciales entre las variables después del análisis factorial, que representa el grado en que los factores «explican» cada uno en los resultados. La diagonal contiene las *medidas de suficiencia de muestreo* para cada variable y los valores fuera de la diagonal son correlaciones parciales entre las variables.

Matriz de correlaciones Tabla que indica las intercorrelaciones entre todas las variables.

Matriz de estructura de factores Una *matriz factorial* que se encuentra en una *rotación oblicua* que representa las correlaciones simples entre las variables y los factores. La mayoría de los investigadores prefieren usar la *matriz de pauta factorial* a la hora de interpretar una solución oblicua.

Matriz de pauta factorial Una de las dos *matrices factoriales* que se encuentra en una *rotación oblicua,* que es lo más parecido a la matriz factorial en una *rotación ortogonal.*

Matriz factorial Tabla que muestra las *cargas de los factores* de todas las variables sobre cada factor.

Matriz sombrero Matriz que contiene valores para cada observación sobre el diagonal conocidos como *valores sombrero,* que representan el impacto de la variable criterio observada sobre el valor de su predicción. Si todos los casos tienen igual influencia, cada uno tendría un valor de (p/n), donde p es igual al número de variables predictor y n es el número de casos. Si un caso no tiene influencia, su valor será $-1/n$, mientras que si un caso ejerce una influencia completamente dominante, el valor será

$(n-1)/n$. Los puntos que exceden dos veces la media $(2p/n)$ se clasifican como *observaciones influyentes*.

Mayor raíz característica (mrc) Estadístico para contrastar la hipótesis nula en el MANOVA. Mide la combinación lineal de las variables dependientes que minimiza las diferencias entre los grupos.

Medida de stress Proporción de la variación de las *disparidades* (datos óptimamente analizados) no tenidos en cuenta por el modelo de análisis multidimensional. Este tipo de medida varía de acuerdo con el tipo de programa y los datos que se están analizando. La medida de stress ayuda a determinar el número apropiado de *dimensiones* incluidas en el modelo.

Medida de suficiencia de muestreo Medida calculada tanto por la matriz de correlación completa como por cada variable individual para evaluar la conveniencia de aplicar el análisis factorial. Valores por encima de 0,50 de la matriz completa o una variable individual indican conveniencia.

Medida del ajuste incremental Medida de la *calidad del ajuste* que compara el modelo con un *modelo nulo* concreto para determinar los grados de mejora sobre el modelo nulo. Complementa las otras dos medidas de calidad del ajuste, las medidas de *ajuste absoluto* y *ajuste de parsimonia*.

Medida extraída de la varianza Cantidad de varianza común o «compartida» entre los *indicadores* o *variables manifiestas* para un *constructo*. Los valores altos representan un mayor grado de representación compartida de los indicadores con el constructo.

Medida multivariante Uso de dos o más variables como *indicadores* de una *medida compuesta* única. Por ejemplo, un contraste de personalidad puede suministrar las respuestas a una serie de preguntas individuales *(indicadores)*, que se combinan para formar un único registro (*escala aditiva*) que representa un rasgo de la personalidad.

Medidas absolutas de ajuste Las medidas de conjunto de la *calidad de ajuste* tanto para los *modelos de medida* como para los *estructurales*. Este tipo de medida no hace ninguna comparación con un *modelo nulo* concreto (*medidas de ajuste incremental*) o ajustar para el número de parámetros en el modelo estimado (*medidas de ajuste de parsimonia*).

Medidas de ajuste de parsimonia Medida de *calidad del ajuste* conjunta que representa el grado de ajuste del modelo por coeficiente estimado. Esta medida intenta corregir cualquier «sobreajuste» del modelo y evalúa la *parsimonia* del modelo comparado con la calidad del ajuste.

Medidas derivadas Procedimiento para obtener percepciones de los encuestados de *datos de similitudes*. Las similitudes derivadas se basan normalmente en series de «puntuaciones» dadas a estímulos por parte de los encuestados que se combinan a continuación en cierta forma. Frecuentemente se utiliza el análisis diferencial semántico para obtener dichas «puntuaciones».

Medidas repetidas Empleo de dos o más respuestas de un solo individuo en un análisis ANOVA o MANOVA. El objetivo del diseño de las medidas repetidas es controlar las diferencias a un nivel individual, que pueden afectar a la varianza dentro de los grupos. Medidas repetidas constituyen una forma de la falta de *independencia* del encuestado.

Metadatos Descripción completa de un elemento de datos, no sólo definiendo su localización y tipo de dato, sino también atributos, dimensiones o asociaciones con cualquier otro elemento de los datos. En términos más sencillos, que se almacenan en un *depósito* de acceso común para todos los usuarios de la base de datos.

Método de aglomeración *Método jerárquico* que empieza con cada *objeto* u observación en unos conglomerados aislados. En cada paso siguiente, los dos objetos que son más parecidos se combinan para construir un nuevo conglomerado agregado. El proceso se repite hasta que todos los objetos se combinan finalmente en un único conglomerado.

Método de comparación pareado Método de presentación de un par de *estímulos* a un encuestado para su evaluación, teniendo el encuestado que elegir un estímulo como el preferido.

Método de composición Aproximación alternativa a los métodos más tradicionales de *descomposición* de elaboración de mapas perceptuales que deriva evaluaciones de *similitud o preferencia* de las evaluaciones de atributos separados por parte de cada encuestado. Estas evaluaciones de atributos separadas se combinan (componen) para una evaluación conjunta. Los ejemplos más habituales de métodos composicionales son las técnicas de análisis de factor y el análisis discriminante.

Método de descomposición Método de elaboración del mapa perceptual asociado con las técnicas de análisis multidimensional en el cual el encuestado sólo responde a una evaluación global de *similitud o preferencia* entre *objetos*. Este conjunto de evaluaciones conjuntas se «descompone» a continuación en un conjunto de «dimensiones» que representan mejor las diferencias de los objetos.

Método de extremos polares El proceso de construir una variable dependiente categórica a partir de una variable métrica. Primero la variable métrica se divide en tres categorías. Después las categorías extremas se utilizan en el análisis discriminante o en la regresión logística, donde la categoría intermedia no está incluida en el análisis.

Método de perfiles completos Método de presentación de *estímulos* a los encuestados para su evaluación, que consiste en una descripción completa de los estímulos para todos los atributos. Por ejemplo, supongamos que un dulce se describe mediante tres factores con dos niveles cada uno: precio (15 o 25 centavos), sabor (cítrico o azúcar y mantequilla) y color (blanco o rojo). Un estímulo de perfil completo estaría definido por un nivel de cada factor. Uno de esos perfiles completos sería un dulce de azúcar y mantequilla rojo que costase 15 centavos.

Método de umbral paralelo Procedimiento *de análisis cluster no jerarquizado* que selecciona simultáneamente las semillas de conglomerados desde el principio. Los *objetos* que caigan dentro de las distancias umbral se asignan a la semilla más cercana. Las distancias umbral pueden ajustarse para que se incluyan más o menos objetos en los conglomerados. Este método es el opuesto del *método de umbral secuencial*.

Método de umbral secuencial Procedimiento *de análisis cluster no jerarquizado* que empieza seleccionando una *semilla de conglomerado*. Todos los *objetos* que estén dentro de una distancia previamente especificada se incluyen en ese conglomerado. Las semillas de conglomerado siguientes se seleccionan hasta que todos los objetos se agrupan en un conglomerado.

Método de Ward Procedimiento de *análisis cluster jerárquico,* en el cual la *similitud* utilizada para unir conglomerados se calcula como la suma de los cuadrados entre los conglomerados para todas variables. Este método tiene tendencia a obtener conglomerados de aproximadamente el mismo tamaño, debido a la minimización de la variación dentro del grupo.

Método del centroide *Algoritmo de aglomeración* en el que la *similitud* entre los agregados se mide como la distancia entre los *centroides de los grupos*.

Método divisivo Procedimiento de aglomeración que empieza con todos los *objetos* en un conglomerado único, que se divide a continuación a cada paso en dos agregados que contienen los objetos más diferentes. Este método es el contrario del *método aglomerativo*.

Método trade-off Método de presentación de estímulos a los encuestados en los que los atributos se representan de dos en dos y los encuestados clasifican todas las combinaciones de los niveles en términos de preferencia.

Métodos de imputación Procesos de estimación de *datos ausentes* basados en valores válidos de otras variables. El objetivo es emplear relaciones conocidas que puedan identificarse en los valores válidos de la muestra para ayudar a la tarea de estimar los datos ausentes.

Mínimos cuadrados Procedimiento de estimación utilizado en la regresión simple y múltiple por lo que se estiman los coeficientes de regresión para aminorar la suma total de los *residuos* cuadrados.

Modelo Conjunto especificado de relaciones de dependencia que puede ser contrastado empíricamente —una operacionalización de una teoría. El propósito de un modelo es proporcionar concisamente una representación amplia de las relaciones a examinar. El modelo puede ser formalizado en un diagrama de secuencias o en un conjunto de relaciones estructurales.

Modelo adaptativo Técnica de simplificación del análisis conjunto que combina los modelos *autoexplicados* y conjunto tradicional. El ejemplo más común es el *análisis conjunto adaptativo* de Sawtooth Software.

Modelo aditivo Modelo basado en la regla de composición aditiva, que asume que los individuos sólo «añaden» los componentes parciales de la utilidad total para calcular un valor de la «utilidad total» o de conjunto que indique la utilidad o la preferencia. También se conoce como un *modelo de efectos principales*.

Modelo autoexplicado Técnica de *composición* para realizar el análisis conjunto en el que los encuestados proporcionan las *estimaciones de los componentes de la utilidad parcial* directamente sin hacer elecciones.

Modelo de composición Clase de modelos univariantes que basan la relación de dependencia en observaciones del encuestado teniendo en cuenta tanto las variables dependientes como las independientes. Tales modelos calculan o «componen» la variable dependiente desde los valores dados por el encuestado para las variables independientes. Los métodos principales son el análisis de regresión y el análisis discriminante. Estos modelos están en contraposición directa con los *modelos de descomposición*.

Modelo de descomposición Clase de modelos multivariantes que «descomponen» la preferencia del encuestado. Esta clase de modelos enfrenta al encuestado con un conjunto predefinido de variables independientes, normalmente en la forma de un producto o servicio hipotético o actual, y a continuación interrogan acerca de una evaluación o preferencia general del producto o servicio. Una vez se ha dado, la preferencia se «descompone» relacionando los atributos conocidos del producto (que se convierten en las variables independientes) con la evaluación (variable dependiente). El más importante de estos modelos es el análisis conjunto y algunas formas de análisis multidimensional (véase Capítulo 10).

Modelo de ecuaciones estructurales (SEM) Técnicas multivariantes que combinan aspectos de la regresión múltiple (examinando relaciones de dependencia) y análisis de factor (que representan conceptos inmedibles —factores— con variables múltiples) para estimar una serie de relaciones de dependencia interrelacionadas simultáneamente.

Modelo de Kohonen Procesamiento secuencial de grandes muestras de observaciones (conocidas como muestras de *entrenamiento*) en las que la predicción o los errores de clasificación se utilizan para recalibrar las ponderaciones para mejorar la estimación.

Modelo de medida Submodelo en SEM que (1) especifica los *indicadores* para cada *constructo*, y (2) evalúa la *fiabilidad* de cada constructo para la estimación de las *relaciones causales*. El modelo de medida es similar en forma al análisis factorial; la principal diferencia reside en el grado de control proporcionado por el investigador. En el análisis factorial, el investigador puede especificar sólo el número de factores, pero todas las variables tienen cargas (es decir, actúan como indicadores) para cada factor. En el modelo de medida, el investigador especifica qué variables son indicadores para cada constructo, con variables que no tienen otras cargas que aquellas de su constructo especificado.

Modelo estructural Conjunto de una o más relaciones de dependencia que vinculan los constructos del modelo supuesto. El modelo estructural es más útil en la representación de las interrelaciones de las variables entre las relaciones de dependencia.

Modelo híbrido Véase *modelo adaptativo.*

Modelo identificado Modelo estructural con cero *grados de libertad* y que cumple exactamente la *condición de orden*. Esto corresponde a un modelo perfectamente ajustado, pero que no tiene generalización.

Modelo infraidentificado *Modelo estructural* con un número negativo de *grados de libertad*. Esto indica un intento por estimar más parámetros que los que son posibles con la matriz de entradas.

Modelo nulo Punto de comparación utilizado en los *índices de ajuste incrementales*. El modelo nulo se supone que es el modelo más simple que puede ser teóricamente justificado. El ejemplo más común es un modelo *constructo* relacionado con todos los *indicadores* sin *error de medida*.

Modelo sobreidentificado Modelo estructural con un número positivo de *grados de libertad*, lo que indica que puede ser posible cierto grado de generalización. El objetivo es alcanzar el ajuste máximo del modelo con el mayor número de grados de libertad.

Modelos anidados Modelos que tienen los mismos constructos pero difieren en términos del número o tipos de *relaciones causales* representadas. La forma más común de modelos anidados se produce cuando una relación se añade o se destruye a partir de otro modelo. Por tanto, el modelo con menores relaciones estimadas está «anidado» dentro del modelo más general.

Modelos equivalentes Modelos comparables que tienen el mismo número de grados de libertad (*modelos anidados*) pero que difieren en una o más secuencias. El número de modelos equivalentes aumenta muy rápidamente a medida que aumenta la complejidad del modelo.

Muestra de análisis El grupo usado para estimar la(s) función(es) discriminante(s) o el modelo de regresión logística. Cuando se construyen las matrices de clasificación, la muestra original se divide aleatoriamente en dos grupos, uno para la estimación del modelo (la muestra de análisis) y el otro para su validación (la muestra ampliada).

Muestra de preparación Observaciones utilizadas para la calibración de una *red neuronal*. Debe contener los valores efectivos para el *nodo* de *ouput*, de tal forma que los errores en el valor de predicción del *output* puedan ser determinados y utilizados en el proceso de *aprendizaje*.

Muestra de validación Véase ampliación de la muestra.

Multicolinealidad Medida en la que puede ser explicada una variable por otras variables del análisis. A medida que aumenta la multicolinealidad, se complica la interpretación del *valor teórico* dado que es más difícil averiguar el efecto de cualquier variable aislada, debido a las interrelaciones entre las variables. Véase *colinealidad*.

Nivel Valor específico que describe a un *factor*. Cada factor debe representarse por dos o más niveles, pero el número de niveles normalmente nunca excede de cuatro o cinco. Si el factor es métrico, debe reducirse a un número pequeño de niveles. Por ejemplo, los muchos posibles valores de tamaño y precio pueden representarse por un reducido número de niveles: tamaño (10, 12 o 16 onzas); o precios (1,19 $, 1,39 $ o 1,99 $). Si la variable no es métrica, se pueden utilizar los valores originales como en los siguientes ejemplos: color (rojo o azul); marca (*X*, *Y* o *Z*); aditivo suavizante de fábrica (presente o ausente).

Nivel de significación Véase *Error de Tipo I*.

Nodo Unidad básica de las *redes neuronales* que pueden actuar como un input, output, o función de análisis/procesamiento. Análogo en su funcionamiento a una neurona en el cerebro.

No-recursiva Relación en un diagrama de secuencias que indica una relación mutua o recíproca entre dos constructos. Cada constructo del par tiene una relación causal con el otro. Representado por una flecha bidireccional entre los dos constructos.

Normalidad Grado en el cual la distribución de los datos de la muestra corresponde a una *distribución normal*.

Objeto Cualquier estímulo, incluido entidades tangibles (productos y objetos físicos), acciones (servicios), percepciones sensoriales (gusto, tacto, vista) o incluso pensamientos (ideas, slogan) que pueden ser comparados y evaluados por el encuestado.

Observación influyente Una observación que tiene una influencia desproporcionada sobre uno o más aspectos de la regresión estimada. Esta influencia puede estar basada en diferencias de respuesta sobre las variables predictor, valores observados extremos para la variable criterio o una combinación de estos efectos. Las observaciones influyentes pueden ser tanto «buenas» al reforzar la pauta de los datos restantes o «malas», cuando un único o pequeño conjunto de casos afecta indebidamente a la regresión estimada. No es necesario que la observación sea un *atípico*, aunque muchas veces los atípicos pueden clasificarse también como observaciones influyentes.

Odds ratio La comparación de la probabilidad de un suceso con la probabilidad de que el suceso no tenga lugar, que se utiliza como variable dependiente en la regresión logística.

Ortogonal Independencia matemática (sin correlación) de los ejes de los factores de cada uno (es decir, ángulos rectos o de noventa grados).

Ortogonalidad Independencia estadística o una ausencia de relación. Los *valores teóricos* ortogonales explican la varianza única, sin ofrecer ninguna explicación de la varianza que es compartida por ellos. Los *contrastes* de ortogonalidad son *comparaciones planificadas* que son estadísticamente independientes y representan comparaciones únicas de las medias de los grupos.

Parámetro Cantidad (medida) característica de la población. Por ejemplo, **(k)** y **(p)** son los símbolos usados para los parámetros de la población media **(k)** y varianza **(p)**. Habitualmente se estiman a partir de los datos de la muestra donde la media aritmética muestral se utiliza como medida de la media poblacional y la varianza muestral se usa para estimar la varianza de la población.

Parrilla de importancia/realización Aproximación bidimensional para ayudar al investigador en la colocación de las dimensiones. El primer eje son las percepciones de importancia de los encuestados (por ejemplo, medida en una escala de «extremadamente importante» o «no importante en absoluto»). El eje opuesto es realización (por ejemplo, medida en una escala de «alta probabilidad de realización» o «alta improbabilidad de realización») para cada marca, servicio o producto, de varios atributos. Cada objeto está representado por sus valores de importancia de realización.

Parsimonia Grado en que un modelo logra la *calidad del ajuste* para cada coeficiente estimado. El objetivo no es minimizar el número de coeficientes o maximizar el ajuste, sino maximizar la cantidad de ajuste por coeficiente es-

timado y evitar «sobreajustar» el modelo con coeficientes adicionales que sólo consigan pequeñas ganancias en el ajuste del modelo.

Perceptor multicapa La forma más habitual de *red neuronal*, que contiene al menos una *capa oculta* de *nodos* ente los nodos *output* y los *input*.

Petición Acceso dirigido a una base de datos requiriendo información específica. Puede utilizarse para la *búsqueda concreta* desde los *datos agregados*.

Polinomiales Transformación de una variable independiente para representar una relación curvilineal con la variable criterio. Incluyendo un término cuadrático (X^2), se estima un único punto de inflexión. Un término cúbico estima un segundo término de inflexión. También se pueden estimar términos adicionales de una potencia mayor.

Ponderación discriminante Ponderación cuyo tamaño se determina por la estructura de varianzas de las variables originales entre los grupos de la variable dependiente. Las variables independientes con gran capacidad discriminante tienen generalmente grandes ponderaciones, y aquellas con escasa capacidad discriminatoria tienen normalmente pequeñas ponderaciones; sin embargo la multicolinealidad entre las variables independientes provocará excepciones a esta regla. También denominado coeficiente discriminante.

Porcentaje correctamente clasificado Véase ratio de aciertos.

Potencia Probabilidad de rechazar correctamente la hipótesis nula cuando es falsa, es decir, encontrar acertadamente una relación supuesta cuando existe. Está definido como 1 - β (véase *beta*). Se determina la potencia como una función de (1) el nivel de significación estadística (α), establecido por el investigador para un *error de Tipo I,* (2) el tamaño muestral utilizado en el análisis, y (3) el *efecto tamaño,* que se examina.

Preferencia Implica que los *objetos* se juzgan por el encuestado en términos de relaciones de dominación; esto es, los estímulos están ordenados en preferencia con respecto a la misma propiedad. La clasificación directa, comparaciones por pares y las escalas de preferencia se utilizan frecuentemente para determinar las preferencias del encuestado.

Preparación Véase *aprendizaje*.

Procedimiento de optimización Procedimiento *de análisis cluster no jerárquico* que permite la reasignación de *objetos* desde el conglomerado originalmente asignado a otro conglomerado sobre la base de un criterio de optimización conjunta.

Procedimientos jerárquicos Procedimientos *cluster* por partes que implican una combinación (o división) de los *objetos* en conglomerados. Estos dos procedimientos son los *métodos aglomerativo* y *divisivo*. El resultado es la construcción de una jerarquía o estructura en árbol (*dendrograma*), que representa la formación de los conglomerados. Este procedimiento arroja $N - 1$ soluciones *cluster*, donde N es el número de objetos. Por ejemplo, si el procedimiento aglomerativo empieza con cinco objetos en distintos conglomerados, mostraría cómo se forman cuatro conglomerados, después tres, después dos y finalmente uno.

Procedimientos no jerárquicos Procedimientos que producen sólo una única solución de conglomerado para un conjunto de semillas de conglomerado. En lugar de utilizar el proceso de construcción en forma de árbol que se encuentra en los *procedimientos jerárquicos,* las *semillas de conglomerado* se utilizan para agrupar *objetos* dentro de una distancia previamente especificada para las semillas. Por ejemplo, si se van a especificar cuatro semillas de conglomerados, sólo se van a formar cuatro conglomerados. Los procedimientos no jerárquicos no producen resultados para todos los posibles conglomerados, tanto como se hace en un *procedimiento jerárquico*.

Procesamiento analítico en línea Aplicación de bases de datos que permite a los usuarios ver, manipular y analizar *datos multidimensionales*.

Proceso de datos ausentes Cualquier evento externo al encuestado (tal como errores en la entrada de datos o problemas en su recogida) o cualquier otra acción por parte del encuestado (como rehusar responder a una pregunta) que lleva a los *datos ausentes*.

Proyecciones Puntos definidos por líneas perpendiculares desde un objeto a un vector. Las proyecciones se utilizan en la determinación del orden de *preferencia* con representaciones de *vector*.

Pseudo R^2 Un valor del ajuste global del modelo que puede calcularse para la regresión logística, comparable a la medida R^2 empleada en la regresión múltiple.

Punto de apalancamiento Una observación se denomina punto de apalancamiento si tiene un impacto sustancial sobre los resultados de regresión debido a sus diferencias respecto del resto de observaciones de una o más variables independientes. La medida más común del punto de apalancamiento es el *valor sombrero,* contenido en la *matriz sombrero*.

Punto ideal Punto sobre un mapa perceptual que representa la combinación más preferida de atributos percibidos (de acuerdo con los encuestados). Un supuesto fundamental es qué posición del punto ideal (en relación con los otros objetos del mapa perceptual) definiría *preferencia* relativa, de tal forma que los objetos más alejados del punto ideal deberían ser menos preferidos.

Puntuación a la inversa Proceso de invertir la puntuación de una variable, a la vez que se mantienen las características distribucionales, para cambiar las relaciones (correlaciones) entre dos variables. Se usa en la construcción de *las escalas aditivas* para evitar una «cancelación mutua» entre las variables con *cargas de los factores* positivas y negativas sobre el mismo factor.

Puntuación de corte El criterio (de puntuación) frente al cual son evaluadas cada una de las puntuaciones Z discriminantes individuales para determinar la pertenencia al grupo predicho. Cuando el investigador incluye dos grupos, la predicción de grupo se determina calculando una única puntuación «de corte». Las entidades cuya puntuación Z está por debajo de ésta son asignadas a un grupo, mientras aquellas cuya puntuación está por encima son clasificadas en otro grupo. Para tres o más grupos, se em-

plean las funciones discriminantes múltiples con una puntuación de corte distinta para cada función.

Puntuación de corte óptima Valor de la puntuación Z discriminante que mejor separa los grupos sobre cada función discriminante con propósitos de clasificación.

Puntuación de factores Medida compuesta creada para cada observación sobre cada factor extraído en el análisis factorial. Se usan las ponderaciones factoriales en relación con los valores de la variable original para calcular la puntuación de cada observación. Así, la puntuación de factores puede ser utilizada para representar el(los) factor(es) en los análisis posteriores. La puntuación de factores está estandarizada para tener una media de 0 y una desviación estándar de 1.

Puntuación *Z* Véase *puntuación Z discriminante.*

Puntuación *Z* discriminante Puntuación referida a la función discriminante para cada objeto en el análisis y habitualmente estandarizada. Denominada también puntuación *Z*, se calcula para cada objeto sobre cada función discriminante, y se utiliza junto a la *puntuación de corte* para determinar la pertenencia al grupo predicho. Es diferente de la terminología de puntuación z utilizada para las variables estandarizadas.

Raíces canónicas *Correlaciones canónicas* al cuadrado, que proporcionan una estimación de la cantidad de varianza compartida entre las respectivas combinaciones lineales *(valor teórico canónico)* óptimamente ponderadas de las variables criterio y de las variables predictoras. También conocidos como autovalores.

Raíz latente Véase *autovalor.*

Ratio de aciertos Porcentaje de objetos (individuos, encuestados, empresas, etc.,) correctamente clasificados por la función discriminante. Se calcula como el número de objetos en la diagonal de la matriz de clasificación dividido por el número total de objetos.

Recíproco Véase *no recursivo.*

Red neuronal Modelo predictivo no lineal que aprende a través del *entrenamiento.* Se parece a la estructura de sistemas neuronales biológicos.

Reespecificación Estimación no estadística de un intervalo de confianza de un parámetro a través del uso de una distribución muestral derivada empíricamente. La distribución empírica se calcula desde múltiples muestras extraídas de la muestra original.

Reespecificación del modelo Modificación de un modelo existente con parámetros estimados para corregir parámetros inapropiados encontrados en el proceso de estimación o para crear un *modelo rival* con el que comparar.

Regla de parada *Algoritmo* para la determinación del número final de conglomerados que se van a formar. Sin una regla de parada inherente al análisis *cluster*, los investigadores han desarrollado varios criterios o pautas generales para su determinación. Existen dos clases de reglas aplicadas post hoc y calculadas por el investigador, que son (1) medidas de *similitud* y (2) medidas estadísticas adaptadas.

Reglas de asociación Reglas basadas en la correlación entre los atributos o *dimensiones* de los *elementos de los datos.*

Reglas de composición Regla utilizada en la combinación de atributos para producir un juicio de valor relativo o *utilidad* para un producto o servicio. A efectos ilustrativos, supongamos que se pregunta a una persona que evalúe cuatro objetos. Se supone que la persona evalúa los atributos de los cuatro objetos y crea cierto valor relativo conjunto para cada uno. La regla puede ser tan simple como crear una ponderación mental para cada atributo percibido y añadir las ponderaciones a una puntuación global (*modelo aditivo*), o puede ser un procedimiento más complejo que implique *efectos interacción.*

Regresión logística Forma especial de regresión en la que la variable dependiente es una variable dicotómica (binaria) no métrica. Aunque existen algunas diferencias, la forma general de interpretación es bastante similar a la de la regresión lineal.

Regresión múltiple Modelo de regresión con dos o más variables independientes.

Regresión simple Modelo de regresión con una única variable independiente.

Regresiones parciales combinando variables Método de seleccionar las variables para incluirlas en el modelo de regresión que considera todas las posibles combinaciones de las variables independientes. Por ejemplo, si el investigador ha especificado cuatro potenciales variables independientes, esta técnica estimaría todos los posibles modelos de regresión con una, dos, tres y cuatro variables. La técnica seleccionaría entonces el modelo con la mejor precisión predictiva.

Relación causal Relación de dependencia entre dos o más variables en las que el investigador claramente especifica que una o más variables «causan» o crean un resultado representado por al menos alguna otra variable. Deben cumplirse los requisitos de la *causación.*

Relación estadística Relación basada en la correlación entre dos o más variables independientes con la variable criterio. Las medidas de asociación, correlaciones típicas, representan el grado de relación porque existe más de un valor de la variable criterio para cada valor de la independiente.

Replicación Repetición de un experimento con el intento de validar los resultados con otra muestra de encuestados.

Representación gráfica multivariante Método de presentación de un perfil multivariante de una observación sobre tres o más variables. Los métodos incluyen aproximaciones tales como *glyphs*, transformaciones matemáticas o incluso representaciones icónicas (por ejemplo rostros).

Residuo Medida del ajuste predictivo para una única observación. Se calcula como la diferencia entre los valores reales y predichos de la variable dependiente. Se supone que los residuos tienen una media de cero y una varianza constante. Juegan un papel importante no sólo en la determinación del cumplimiento de los supuestos subyacentes de regresión, sino también sirven como una herramienta de diagnosis en la identificación de *casos atípicos* y *observaciones influyentes.*

Residuo eliminado Proceso de calcular los *residuos* en los cuales se elimina la influencia de cada observación cuan-

do se calcula su residuo. Esto se consigue mediante la *i*-ésima observación de la ecuación de regresión que se usa para calcular su valor de predicción.

Residuo estandarizado El reajuste del *residuo* hacia una base común con la división de cada residuo por la desviación estándar de los residuos. Por tanto, los residuos estandarizados tienen una media de 0 y una desviación estándar de 1. Cada valor de residuo estandarizado puede ser considerado en términos de errores estándar en cuanto a los tamaños muestrales medianos hasta grandes. Esto proporciona una manera directa de identificar los casos atípicos al igual que con los valores por encima de 1 o 2 para niveles de confianza de 0,10 y 0,05, respectivamente.

Residuos studentizados La forma más común de *residuos estandarizados*, que difiere del resto de los métodos en cómo calcula la desviación estándar utilizada en la estandarización. Para minimizar el efecto de un *caso atípico* aislado, la desviación del residuo estándar de una observación, utilizado para estandarizar el *i*-ésima residual, se calcula desde los estimadores de la regresión omitiendo la *i*-ésima observación. Esto se realiza repetidamente para cada observación, omitiendo cada vez una observación del cálculo. Este enfoque es similar al *residuo eliminado,* aunque en esta situación la observación es omitida del cálculo de la desviación estándar.

Retropropagación El proceso de *aprendizaje* más habitual en las *redes neuronales*, en el que los errores de estimación de los *nodos* del *output* se a través del sistema y se utilizan como indicadores de cómo calibrar las ponderaciones de cada nodo.

Rotación de factores Proceso de manipulación o ajuste de los ejes del factor para lograr una solución factorial más sencilla y básicamente más significativa.

Rotación factorial oblicua *Rotación de factores* calculada para que los factores extraídos estén correlacionados. En vez de obligar arbitrariamente a la rotación de factores a una solución *ortogonal,* la rotación oblicua identifica el grado en que cada uno de los factores se correlacionan.

Rotación ortogonal factorial Rotación de factores en la que los factores son extraídos de tal forma que sus ejes se mantienen a 90 grados.

SDFBETA Véase *DFBETA.*

SDFFIT Véase *DFFIT.*

Semillas de conglomerados *Centroides* iniciales o puntos de partida de los conglomerados. Estos valores se seleccionan para empezar los *procedimientos no jerárquicos del análisis cluster*, en el que los conglomerados se construyen alrededor de estos puntos previamente especificados.

Sesgo de nodo *Nodo* adicional de una *capa oculta*, que tiene un valor constante. Este nodo funciona de forma similar al término constante de una regresión múltiple.

Significación práctica Medios para valorar en el análisis multivariante resultados basados en hallazgos considerados sustantivos antes que en su significación estadística. Mientras la significación estadística parece indicar que el resultado es atribuible a la casualidad, la significación práctica juzga si el resultado es útil (esto es, suficiente como para asegurar la acción).

Simetría Medida de la simetría de una distribución; en muchos casos la comparación se hace para una *distribución normal.* Una distribución simétrica positiva contiene pocos valores elevados y se inclina hacia la derecha, y una distribución simétrica negativa tiene pocos valores bajos y se inclina hacia la izquierda. Cuando los valores de simetría caen fuera del rango de –1 a +1 indican una distribución sustancialmente simétrica.

Similitud Véase *datos de similitudes.*

Similitud entre objetos La correspondencia o asociación de dos *objetos* basada en las variables del *valor teórico del análisis* cluster. La similitud puede medirse de dos formas. La primera es una medida de asociación, donde los coeficientes de correlación positivos más elevados representan la mayor similitud. La segunda, «proximidad» o «cercanía», entre cada par de objetos puede evaluar la similitud, donde se utilizan las medidas de distancia o diferencia, y donde las distancias o diferencias pequeñas representan la mayor similitud.

Simulación Creación de múltiples matrices de entrada de datos basándose en parámetros específicos que reflejan la variación en la distribución de la entrada de datos.

Simulador de elecciones Procedimiento que permite al investigador evaluar muchos escenarios del tipo «qué pasa si», incluyendo la preferencia por posibles configuraciones de productos o servicios o las interacciones competitivas entre los estímulos que se supone constituyen un mercado. Una vez que se han estimado los *componentes parciales de la utilidad total* conjuntos para cada encuestado, el simulador de elección analiza un conjunto de *estímulos de perfil completo* y predice tanto elecciones individuales como agregadas para cada estímulo del conjunto.

Singularidad El caso extremo de *colinealidad* o *multicolinealidad* en el que se puede predecir una variable independiente perfectamente (una correlación de 6 1,0) por una o más variables independientes. No se pueden estimar los modelos de regresión cuando existe la singularidad. El investigador tiene que omitir una o más de las variables involucradas para quitar la singularidad.

Sistemas de apoyo a la decisión (DSS) Sistemas interactivos desarrollados para proporcionar a los usuarios datos de (1) preguntas ad-hoc o mal definidas, (2) informes preespecificados (es decir, informe de excepción) y (3) análisis estructurado (procedimientos basados en reglas, estadísticas multivariantes, redes neuronales y otros modelos).

Sistemas operacionales Sistemas de información que proveen las funciones básicas de la organización (contabilidad, gestión de existencias, procesamiento de pedidos) para sus operaciones del día a día.

Solución degenerada Solución de análisis multidimensional que no es válida porque (1) hay inconsistencias en los datos o (2) muy pocos objetos comparado con la dimensionalidad de la solución. Incluso aunque el programa informático pueda indicar una solución válida, el investigador no debería tener en cuenta la solución degenerada y examinar los datos para encontrar la causa. Este tipo de soluciones se representa habitualmente como un argumento circular con resultados ilógicos.

Suma del cuadrado de la regresión (SSR) Suma de las diferencias al cuadrado entre la media y los valores de la predicción de la variable independiente para todas las observaciones. Representa la cantidad de mejora en la explicación de la variable independiente atribuible a las variables independientes.

Suma del cuadrado de los errores (SSE) Suma de los errores de *predicción al cuadrado (residuos)* a lo largo de todas las observaciones. Usada para denominar a la varianza en las variables dependientes no tenidas todavía en cuenta en el modelo de regresión. Si no se utilizan variables independientes para la predicción, se convierte en los errores cuadrados usando la media como el valor predictor y por tanto iguala a la *suma total de los cuadrados*.

Suma total de los cuadrados (TSS) Cantidad total de variación que existe para ser explicada por las variables independientes. Esta «línea de fondo» se calcula sumando las diferencias al cuadrado entre la media y los valores efectivos de la variable criterio para todas las observaciones.

Supervisión Proceso de *aprendizaje* que emplea una *muestra de preparación* y proporciona retroalimentación a la *red neuronal* en lo que se refiere a los errores de los *nodos* de resultado.

T2 de Hotelling Contraste para valorar la significación estadística de la diferencia entre dos conjuntos de medias muestrales. Es un caso especial del MANOVA empleado con dos grupos o niveles de una variable de tratamiento.

Tabla de contingencia Tabulación cruzada de dos variables categóricas o no métricas en la que las entradas son las frecuencias de respuestas que caen en cada «celda» de la matriz. Por ejemplo, si tres marcas se califican con cuatro atributos, la tabla de contingencia de marca por atributo tendría tres filas y cuatro columnas. Las entradas serían el número de veces que una marca se califica como teniendo un atributo.

Tabla de tabulación cruzada Véase *tabla de contingencia*.

Taxonomía Clasificación derivada empíricamente de *objetos* reales basada en una o más características. Tipificada por la aplicación del análisis *cluster* u otros procedimientos de agrupación. Se puede contrastar esta clasificación con una *tipología*.

Teoría Un conjunto sistemático de *relaciones causales* que proporciona una explicación consistente y exhaustiva de un fenómeno. En la práctica, una teoría es un intento de un investigador de especificar el conjunto completo de relaciones de dependencia que explican un conjunto particular de resultados. Una teoría puede basarse en ideas generadas por una o más de las tres fuentes principales de éstas: (1) investigación empírica previa; (2) experiencias pasadas y observaciones de actitudes, comportamientos u otros fenómenos reales; y (3) otras teorías que ofrecen una perspectiva para el análisis. Por tanto, la construcción de teorías no es un dominio exclusivo de los investigadores académicos; también tienen un papel explícito los investigadores aplicados. Para cualquier investigador, la teoría ofrece un medio de dibujar el «gran cuadro» y evaluar la importancia relativa de varios conceptos en una serie de relaciones.

Tipificación Proceso por el cual los datos iniciales se transforman en nuevas variables de medición con una media 0 y una desviación estándar de 1. Cuando los datos se transforman de esta manera, el término b_0 (la constante) supone un valor 0. Cuando se usan datos tipificados, los coeficientes de regresión se conocen como *coeficientes beta*, que permiten al investigador comparar el efecto relativo de cada variable independiente sobre la variable criterio.

Tipificación centrada por filas Véase *tipificación entre sujetos*.

Tipificación intra sujeto Método de estandarización en el cual las respuestas de un encuestado se comparan para la muestra conjunta sino para sus propias respuestas. También conocido como «ipsitizing», las respuestas medidas del encuestado se utilizan para tipificar sus propias respuestas.

Tipología Clasificación conceptual de *objetos* basada en una o más características. Una tipología no intenta normalmente agrupar observaciones efectivas, sino que proporciona los fundamentos teóricos de la creación de una *taxonomía*, que agrupa las observaciones efectivas.

Tolerancia Proporción de la variación de las variables independientes no explicada por las variables ya incluidas en el modelo (función). Se puede usar como protección frente a la multicolinealidad. Se calcula como 1 - R_i^{2*}, donde R_i^{2*} es la parte de varianza de la variable independiente i explicada por el resto de las variables independientes. Una tolerancia de 0 significa que la variable independiente considerada es una combinación lineal perfecta de variables que ya están en el modelo (ecuación). Una tolerancia de 1 significa que una variable independiente es totalmente independiente de las otras ya incluidas en el modelo.

Transacción Acción sobre un *elemento operativo de datos* (creación, modificación o destrucción) que representan un único evento del negocio. Los ejemplos incluyen la venta de un producto a un cliente, la factura de un pago, la transferencia de fondos entre cuentas o la transferencia de productos comprados o terminados a existencias.

Transformación Una variable puede tener una característica indeseable, tal como la no-normalidad, que disminuye la capacidad del coeficiente de correlación para representar la relación entre ésta y cualquier otra variable. Una transformación, tal como tomar el logaritmo o la raíz cuadrada de la variable, crea una nueva variable y elimina la característica indeseable, permitiendo así una mejor medida de la relación. Las transformaciones pueden aplicarse tanto a las variables dependientes como independientes o a ambas. El tipo específico y necesario de transformación puede basarse en razones teóricas (tales como transformar una relación no lineal conocida) o por razones empíricas (identificadas por medios gráficos o estadísticos).

Transformación logit Transformación de los valores de la variable dependiente binaria discreta de la regresión logística en una curva en forma de S (curva logística) que representa la probabilidad de un suceso. Esta probabilidad

se utiliza para formar el odds ratio, que actúa como la variable dependiente en la regresión logística.

Transformaciones de datos Una variable puede tener una característica indeseable, tal como la no normalidad, que desaconseja su uso en la técnica multivariante. Una transformación, tal como tomar el logaritmo o la raíz cuadrada de la variable, crea una variable transformada que es más adecuada para representar la relación. Las transformaciones pueden aplicarse tanto a la variable dependiente como a la independiente o a ambas. El tipo específico y necesario de transformación puede escogerse atendiendo a razones teóricas (tal como transformar un conocida relación no lineal) o razones empíricas (identificada a través de medios gráficos o estadísticos).

Tratamiento Variable independiente que el investigador altera para ver los efectos (si existen) sobre las variables dependientes. La variable de tratamiento puede tener varios niveles. Por ejemplo, diferentes intensidades de las pretensiones en los anuncios publicitarios podrían ser manipuladas para ver el efecto sobre la credibilidad de los consumidores.

Traza Representa la cantidad total de varianza sobre la que se basa la solución de factores. La traza es igual al número de variables, basada en la suposición de que la varianza en cada variable es igual a 1.

Unidimensionalidad Característica de un conjunto de *indicadores* que sólo tienen un rasgo o concepto subyacente en común. Del ajuste entre los indicadores escogidos y la definición teórica del *constructo unidimensional*, el investigador debe establecer tanto conceptual como empíricamente que los indicadores son medidas fiables y válidas de sólo el constructo especificado antes de establecer la unidimensionalidad. Similar al concepto de *fiabilidad*.

Utilidad Un juicio de preferencia subjetivo hecho por un individuo que representa el valor holístico o importancia de un objeto específico. En el análisis conjunto, la utilidad se supone que se forma por la combinación de las estimaciones de los *componentes parciales de la utilidad total* para cualquier conjunto especificado de *niveles* con el uso de un *modelo aditivo*, quizá en conjunción con *efectos interacción*.

Validación aparente Véase *validación de contenido.*

Validación con división de la muestra Véase *validación cruzada.*

Validación cruzada Proceso de división de la muestra en dos partes: la muestra de análisis usada para estimar la(s) función(es) discriminante(s) y la ampliación de la muestra empleada para validar los resultados. La validación cruzada evita el «sobreajuste» de la función discriminante o de la regresión logística permitiendo su validación con una muestra totalmente separada.

Validación de contenido Valoración del grado de correspondencia entre los artículos seleccionados para constituir una *escala aditiva* y su *definición conceptual.*

Validación de criterios Capacidad de los conglomerados para mostrar las diferencias esperadas en una variable no utilizada para formar los conglomerados. Por ejemplo, si los conglomerados se forman sobre calificaciones de realización, el marketing sugeriría que los conglomerados con mayores calificaciones de realización tendrían también las mayores puntuaciones de satisfacción. Si esto fuera así en la contrastación empírica, entonces se apoya al criterio de validación.

Validez Grado en que una medida o conjunto de medidas representan correctamente el concepto que se estudia, es decir, el grado en que se está libre de errores sistemáticos o no aleatorios. La validez se refiere a la bondad con que la(s) medida(s) definen el concepto, mientras que la *fiabilidad* se relaciona con la coherencia de la(s) medida(s).

Validez predictiva Véase *validación de criterios.*

Valor crítico Valor de un contraste estadístico (contraste *t*, contraste *F)* que denota un nivel de significación especificado. Por ejemplo, 1,96 denota un nivel de significación de 0,05 para el contraste *t* con tamaños muestrales grandes.

Valor de la verosimilitud Medida utilizada en la regresión logística o en la regresión logit para representar la ausencia de ajuste predictivo. Incluso aunque estos modelos no utilicen el procedimiento de los mínimos cuadrados en la estimación del modelo, como se hace en la regresión múltiple, el valor de la verosimilitud es similar a la suma del error al cuadrado del análisis de regresión.

Valor de partida Cálculo del parámetro inicial utilizado para los procesos de estimación incrementales o iterativos, como el LISREL.

Valor sombrero Véase *matriz sombrero.*

Valor teórico Combinación lineal de variables compuesta a partir de ponderaciones empíricas aplicadas a un conjunto de variables determinadas por el investigador.

Valores *F* (o *t*) parciales El test parcial *F* es simplemente un test *F* para la contribución adicional de la precisión predictiva de una variable añadida a la de las variables ya presentes en la ecuación. Cuando una variable (X_a) se añade a una ecuación de regresión después de que muchas otras variables ya se han introducido en la ecuación, su contribución puede ser muy pequeña. La razón es que X_a está altamente correlacionada con las variables ya presentes en la ecuación. Un valor parcial *F* se puede calcular para todas las variables pretendiendo simplemente que cada una de las variables sea la última en entrar en la ecuación. Este método arroja la contribución adicional de cada variable por encima de las otras de la ecuación. Un valor *t* puede calcularse en lugar de con valores *F* en todos los casos, con el valor *t* que es la raíz cuadrada del valor *F*.

Valores teóricos canónicos Combinaciones lineales que representan la suma ponderada de dos o más variables y que pueden ser definidos tanto para variables criterio como para variables predictoras. También conocido como *Combinaciones lineales,* compuestos lineales y asociaciones lineales.

Variable categórica Variable que emplea valores que sirven meramente como etiqueta o medio de identificación. También referida como variable no métrica, nominal, binaria, cualitativa o taxonómica. Un ejemplo es el número de la camiseta de un jugador de fútbol. En el Capítulo 1 se encuentra una discusión más completa de sus características y diferencias respecto a una variable métrica.

Variable criterio (*Y*) Variable a predecir o explicar por el conjunto de variables independientes.

Variable dependiente Resultado supuesto de, en respuesta a, un cambio en las *variables independientes.*

Variable ficticia Variable independiente usada para explicar el efecto que tienen los diferentes niveles de una variable no-métrica en la predicción de la variable criterio. Para explicar los L niveles de una variable independiente, se necesitan $L - 1$ variables ficticias. Por ejemplo, el sexo se mide como masculino o femenino y podría ser representado por dos variables ficticias (X_1 y X_2). Cuando el encuestado es hombre, $X_1 = 0$ y $X_2 = 0$. Asimismo, cuando el encuestado es mujer, $X_1 = 0$ y $X_2 = 1$. Asimismo, cuando $X_1 = 1$, sabemos que X_2 debe ser igual a 0. Por tanto necesitamos sólo una variable, X_1 o X_2, para representar la variable género. No tenemos que incluir las dos variables, porque una de ellas se predice perfectamente por la otra (una *singularidad)* y no se puede estimar los coeficientes de regresión. Si una variable tiene tres niveles, sólo se necesitan dos variables ficticias. Por tanto, el número de variables ficticias es uno menos que el número de niveles de la variable no métrica menos uno.

Variable independiente Variable o variables elegidas como predictores y variables que potencialmente explicarían la variable criterio.

Variable manifiesta Valor observado para un ítem o cuestión específica, obtenida bien de los encuestados en respuesta a cuestiones (como en un cuestionario) o de observaciones del investigador. Las variables manifiestas se utilizan como *indicadores* de *constructos* o *variables latentes.*

Variable métrica Variable con una unidad constante de medida. Si una variable es escalada de 1 a 9, la diferencia entre 1 y 2 es la misma que entre 8 y 9. Una discusión más completa de sus características y diferencias respecto a una variable categórica se encuentra en el Capítulo 1.

Variable o concepto no observado Véase *variable o constructo latente.*

Variable o constructo latente Operacionalización de un *constructo* en un modelo de ecuaciones estructurales. Una variable latente no puede ser medida directamente pero puede ser representada o medida por una o más variables (*indicadores*). Por ejemplo, la actitud de una persona hacia un producto nunca puede ser medida con tanta precisión como para que no exista incertidumbre, pero preguntándole varias cuestiones podemos evaluar los diversos aspectos de la actitud de una persona. Combinadas, las contestaciones a esas cuestiones dan una medida razonablemente precisa del constructo latente (actitud) para un individuo.

Variable predictor (X_n) Véase *variable independiente.*

Variable suplente Selección de una variable simple con la *carga de los factores* para representar un factor en la etapa de reducción de datos en vez de usar *una escala aditiva* o una *puntuación de factores.*

Variables criterio (*Y*) Véase *variables dependientes.*

Varianza compartida Cantidad total de varianza que comparte una variable original con todas las demás variables que se incluyen en el análisis.

Varianza del coeficiente de regresión-matriz de descomposición Método de determinación de la contribución relativa de cada *autovalor* para cada coeficiente estimado. Si dos o más autovalores están altamente asociados con un coeficiente aislado, se indica un aceptable nivel de *multicolinealidad.*

Varianza del error La varianza de una variable debido a los errores en la recolección de datos o medición.

Varianza específica Varianza de cada variable única para esta variable y sin explicación o asociación con otras variables en el análisis factorial.

VARIMAX Uno de los métodos de rotación de factores ortogonal más populares.

Vector Conjunto de números reales (por ejemplo., $X_1 \ldots X_n$). Los vectores pueden ser escritos tanto en filas como en columnas. Se consideran los vectores columna como la forma convencional mientras que las vectores fila son considerados como traspuestos. A continuación se representa a los vectores columna y los vectores fila.

Visualización de datos Técnicas de representación de los datos en formato gráfico, utilizados habitualmente a modo de exploración para identificar las relaciones básicas.

Índice analítico

A

Adición progresiva, 143, 171-172
Alfa de Cronbach, 79, 106, 612, 638
Algoritmo genético, 707, 720
Algoritmo, 491
Almacenamiento de datos, 707, 709-714
ALSCAL, 557
AMOS, 634
Ampliación de la muestra: véase también Validación de la muestra
  en el análisis conjunto, 435
  en el análisis discriminante múltiple, 250, 263
ANACOR, 531
Análisis agregado
  en el análisis conjunto, 416, 435-436, 561, 563
  en el análisis multidimensional, 547, 555
Análisis Cluster, 12-13, 79, 85, 491-532
  algoritmos, 508, 510-515
  aplicaciones, 492-493
  coeficiente de aglomeración, 522-523
  detección de atípicos, 500-501, 520-525
  diseño de la investigación, 500-507
  ejemplo de, 518-531
  elección entre jerárquico y no jerárquico, 514-515
  interpretación, 517
  medidas de semejanza 495, 501-507
  número de clusters, 516
  objetivos, 498-500
  obtención de clusters, 495-496
  perfil gráfico de, 527
  procedimientos jerárquicos, 495, 510-514
  procedimientos no jerárquicos, 513-514, 521
  proceso de decisión, 497-498
  re-especificación, 526
  regla de parada, 516
  representación gráfica de los resultados, 524
  selección de variables, 498-500
  supuestos, 507-508
  tipificación de datos507-508
  uso de, 493-497
  validación, 517-518
Análisis conjunto, 12, 407-453
  análisis agregado frente a desagregado, 416, 435-436
  análisis de rentabilidad con, 437
  aplicaciones, 414
  cálculo de valoraciones parciales, 447-448
  comparación con otros métodos multivariantes, 415-416
  diseño de estímulos factoriales frente a fraccionales, 430-432
  diseño de estimulos, 421-424
  ejemplos, 409-413, 440-441
  enfoque basado en la elección, 440-443
  ilustración de, 443-450
  importancia de los atributos, 436
  interpretación, 436, 446, 448
  métodos de presentación, 427-429
  objetivos, 587-591
  proceso de decisición, 416-418
  recogida de datos, 427-433, 445-446
  reglas de composición, 424-426
  selección de medidas (dependientes) de preferencia, 425-426, 430-432
  simulador de elección, 437-438, 449-450
  supuestos, 433, 446
  técnicas de estimación, 433-435
  tipos de relaciones, 416, 426-428
  tratamiento de un elevado número de factores, 439-443
  validación, 436-437
  valoración parcial, 408, 427-428, 447-448
Análisis de clasificación: véase Análisis cluster
Análisis de componentes principales: véase Análisis de factores; Análisis de componentes
Análisis de componentes, 79, 89, 91-92
Análisis de correlación canónica (CCA, 12, 469-487
  aplicaciones, 472
  cargas cruzadas, 479, 484-485
  cargas, 479
  ejemplo, 470-471, 480-486
  estadísticos de contrastes multivariantes, 476, 481
  interpretación de las funciones canónicas, 478-479

objetivos, 473
obtención de funciones canónicas, 476-478
ortogonalidad de funciones, 476
ponderaciones, 478, 484
proceso de decisión, 473-474
redundancia, 477-478, 481-482, 483
significación estadística, 481
supuestos, 474
validación, 479-480
Análisis de correspondencia, 13, 547, 571-578
calculando una medida de asociación, 572-574
diseño de la investigación, 575-576
ejemplo, 588-592
interpretación, 576-577
número de dimensiones, 577
objetivos, 571-575
supuestos, 576
validación, 577
Análisis de correspondencias múltiples: véase Análisis de correspondencias
Análisis de corroboración, 612, 613
Análisis de covariaciones teóricas: véase Análisis multivariante de la varianza
Análisis de datos, cambios de, 708-709
Análisis de factores comunes, 11, 79, 89, 91-92
Análisis de factores corroborantes: véase Modelo de ecuaciones estructurales
Análisis de factores, 11, 79-124
aplicaciones, 103-107
cargas factoriales, 95
comparación con el análisis cluster, 84
definición conceptual, 105
denominación de factores, 102
diseño de, 86-88, 108
ejemplos de, 82-83, 107-121
factor común frente a análisis de componentes, 11, 89, 91-92
interpretación de factores, 95-102
interpretación de la matriz factorial, 100-102
medidas de adecuación muestral, 88-89
número de factores, 92-95
objetivos, 84-86
obtención de factores, 89
proceso de decisión, 82-107
puntuación al revés, 117-118
puntuaciones de factores, 106
requisitos de tamaño muestral, 87-88
rotación de factores, 95-99, 112-113
selección de variables sustituidas, 104, 116
selección de variables, 85-86
selección del método, 106-107
significación estadística de las cargas, 99
significación práctica de cargas, 99
supuestos, 88-89, 108-110
validación, 102-103, 115
Análisis de la estructura de covarianzas: véase Modelo de ecuaciones estructurales
Análisis de la varianza (ANOVA), 2, 348-349
cálculo de efectos, 348-349
ejemplos, 372-382
muestra de análisis, el análisis discriminante múltiple, 250, 363
Análisis de reducción, 215, 369-370, 396
Análisis de regresión: véase Regresión múltiple
Análisis de relaciones, 612
Análisis del factor «Q», 80, 84
Análisis desagregado
en el análisis conjunto, 415-416, 435-436
en el análisis multidimensional, 547, 555
Análisis discriminante múltiple (MDA), 11, 249-322
analogía con la regresión y ANOVA, 252
cambio de intenciones y, 256-259
cargas discriminantes, 250, 276
con análisis multivariante de la varianza (MANOVA), 252
división de la muestra, 263-264, 285
ejemplos, 252-259, 284-299
ejemplos, grupo de dos, 284-299
ejemplos, grupo de tres, 299-316
estimación de la función discriminante, 287, 289
evaluación de las diferencias de grupo, 267-268
frente a análisis multivariante de la varianza (MANOVA), 351
índice de potencia, 277-279, 312-313
interpretación de resultados, 276-279, 296-297
matrices de clasificación, 268-274, 292-293, 305-306
métodos de cálculo, 264-265
objetivos, 261
obtención de funciones discriminantes, 257-258, 301-304
ponderaciones discriminantes Z, 267
precisión al clasificar, 273-274, 292-296, 305-307
probabilidades de clasificación, 268-274
proceso de decisión, 260-280
Q de Press, 274, 292, 307
representación geométrica, 255-256
representación gráfica de resultados, 278, 313-315
requisitos de tamaño muestral, 263, 285
rotación de funciones, 277
selección de variables, 262, 285
significación estadística, 266, 303, 304
supuestos, 264-265, 285-286, 299-300
validación de los resultados, 279-280
valores parciales F, 276-277
y regresión logística, 316-322
Análisis discriminante: véase Análisis discriminante múltiple
Análisis exploratorio, 612, 613
Análisis externo (MDS), 567
Análisis factorial R, 80, 84
Análisis interno (MDS), 567
Análisis *jackknife,* 279, 612, 634, 730
Análisis logit: véase Modelo de probabilidad lineal
Análisis métrico (MDS), 558
Análisis multidimensional (MDS), 13, 547-594
análisis agregado frente a desagregado, 555
análisis de correspondencia, 571-578, 588-592
análisis interno frente a externo, 567

aplicaciones, 547-549
comparación con otros métodos multivariantes, 551-552
datos de semejanza frente a datos de preferencias, 554
determinación de la dimensión, 563-656
dimensiones percibidas frente a objetivas, 548
diseño de la investigación, 566-560
ejemplo de, 578-593
ejemplo hipotético, 549-551
incorporación de datos sobre las preferencias, 565-569, 579, 582-584
índice de ajuste, 547, 564
interpretación de las dimensiones, 569-571, 586-587
medida de tensión, 564
medidas de semejanza, 558-559
métodos de composición, 557-558
métodos de descomposición, 556-558
métodos métricos frente a no métricos, 558
objetivos, 562-555
proceso de decisión, 552-571
puntos ideales, 565-568
recogida de datos, 558-560
representación puntual de puntos ideales, 568-569
representación vectorial de atributos, 569, 586
representación vectorial de puntos ideales, 568-569
selección de objetos de análisis, 554, 558
solución degenerada, 547, 563
supuestos, 560-561
validación, 571, 592-593
Análisis multivariante de la varianza (MANOVA), 2, 3-24, 12, 345-398
análisis de reducción, 369-370, 395, 396
comparaciones planificadas, 371
con el análisis discriminante múltiple (MDA), 252
constrastes post hoc, 369-371
contrastes de significación, 366
contrastes estadísticos de las variables dependientes, 369-370
contrastes multivariantes, 367-369, 393, 394
contrastes univariantes (ANOVA), 348-349
control del porcentaje de error experimental, 354
covariaciones teóricos (MANCOVA), 346, 361-362, 368-369
diseño de la investigación, 357-362
diseño factorial de tratamientos, 357-358
efectos de interacción, 358-360, 394
ejemplo de ordenador, diseño factorial, 388-395
ejemplo, dos grupos, 372-380
ejemplo, grupos independientes k, 380-388
ejemplo, tres grupos, 383-388
factor en bloques, 358
frente a análisis discriminante múltiple (MDA), 351
ilustración, 351-353
interpretación, 368-371
medidas repetidas, 362
objetivos, 356-357
potencia367-368
proceso de decisión, 354-372
replicación, 371-372
requisitos de tamaño muestral, 357
supuestos, 362-364
validación, 371-372
variables dependientes, 354
Análisis Q: véase Análisis cluster
Análisis univariante de la varianza: véase Análisis de la varianza
Análisis univariante, 12
Analistas de datos, 3
Apoyo, 707, 718
Aprendizaje, 707
en redes neurales, 722-723
Árbol de decisión, 707
Árboles de decisión, 718-719
Atípico, 31, 57-63, 207, 210
descripción y perfil, 59
detección bivariante, 58, 61
detección multivariante, 58-59, 61
detección univariante, 58
en el análisis cluster, 500-501, 510-515
en el análisis multivariante de la varianza, 177-179, 364
en la regresión múltiple, 143
ilustración de, 59-73
tratamiento, 59
Ausencia aleatoria (MAR), 31-32, 43
Ausencia totalmente aleatoria (MCAR), 32, 43
Autovalor
en al regresión múltiple, 207, 208
en el análisis factorial, 80, 92-93
en la correlación canónica, 469, 476

B

Bases de datos para ejemplos, 24-27
Beta, 2, 16-20, 346
Bondad de ajuste (SEM), 612, 632, 637-641, 680-690
ajuste conjunto del modelo, 637-638
ajuste del modelo de medida, 632, 640, 649-650
ajuste del modelo estructural, 640
criterio de información de Akaike (AIC), 685
chi cuadrado normada, 684-685
ejemplo de ordenador, análisis de factores corroborante, 646-649
ejemplo, modelo estructural, 689
error de aproximación cuadrático medio (RMSEA), 682
estadístico de verosimilitud chi cuadrado, 680-681
índice de ajuste normado (NFI), 683
índice de ajuste normado parsiminioso (PNFI), 684
índice de bondad de ajuste parsiminioso (PGFI), 684
índice de bondad del ajuste (GFI), 681
índice de bondad del ajuste corregido (AGFI), 683
índice de Tucker-Lewis (TLI), 683
índice de validación cruzada esperada (ECVI), 682
parámetro de no centralidad escalado, 681
parámetro de no centralidad, 681
residuo cuadrático medio (RMSR), 682
Bootstrapping, 612, 633, 730, 731-733
Box plot
definición, 31, 35-36
uso, 375, 383, 386, 388, 392
Búsqueda concreta, 707, 714, 717

C

Capa oculta, 707, 722
Cargas canónicas, 469, 479
Cargas cruzadas canónicas, 469, 479, 484-485
Cargas cruzadas véase Cargas cruzadas canónicas
Cargas discriminantes, 250, 276
Cargas factoriales, 79, 95
CART, 718
Casos *Heywood,* 612, 637
Categoría de referencia, 143
Causación, 612, 620
Centroide de cluster, 491
Centroides, 151, 267
  en el análisis de conglomerados, 491
  en el análisis discriminante, 250
  trazo de, 291
Cluster de K-medios, 513
Clusters subjetivos: véase Datos confundidos
Codificación de efectos, 31, 76-77, 162
  en la regresión múltiple, 143
Codificación de indicador, 31, 76, 143, 162
Codificar, 707, 722
Coeficiente beta, 143, 180-181
Coeficiente de aglomeración, 522-523
Coeficiente de correlación de Pearson, 77
Coeficiente de correlación parcial 143, 171, 182, 191
Coeficiente de correlación partido, 143
Coeficiente de correlación, 143, 146-147
Coeficiente de determinación (R^2)
  cálculo, 151
  coeficiente de determinación corregido, 174-175
  definición, 143
  significación estadística, 174-175
  uso, 191
Coeficiente de determinación corregido, 143, 174-175
Coeficiente de regresión 143, 148
Coeficiente discriminante, 250
Coeficiente logístico, 250, 282
Colinealidad: véase también Multicolinealidad, 143, 151, 207
Combinaciones lineales: véase Valor teórico
Comparaciones planificadas, 346, 371
Concepto (SEM), 612
  endógeno, 613, 624
  exógeno, 613, 624
Conceptos no observados, 612
Condición de rango, 612, 636
Conector completo, 491, 511-512
Conector medio, 491, 513
CONEXIÓN NEURAL, 727
Configuración inicial (MDS), 561, 563
Conjunto de elección, 408, 440
Constante, 143, 147
Constructo endógeno, 612, 624
Constructo exógeno, 612, 614
Constructo/variable latente, 532
Contraste, 346, 371
Contraste a priori, 346, 369
Contraste de caída, 93, 110, 111
Contraste de constrastes de Scheffe, 370
Contraste de esfericidad de Bartlett, 79, 88
Contraste de Levene, 67
Contraste de Newman-Kuels, 370
Contraste de rango múltiple de Duncan, 370
Contraste F: véase ratio F
Contraste global: véase Contraste de comparación múltiple
Contraste M de Box, 250
  en el análisis discriminante múltiple, 286
  en el análisis multivariante de la varianza, 346, 363
Contraste reiterativo: véase Contraste de comparación múltiple
Contraste t, 346, 347-348, 373-375
Contrastes conjuntos: véase Contrastes de comparación múltiple
Contrastes de comparación múltiple
  constrastes post hoc, 346, 369-371
  contrastes a priori, 369, 371
  desigualdad de Bonferroni, 369
Contrastes de comparación: Véase Contrastes de comparación múltiple
Contrastes post hoc: véase Contrastes de comparación múltiple
CORRAN, 576
Correlación biserial, 612, 631
Correlación inter-atributos, 408, 422
Correlación parcial bivariante, 2, 12
Correlación policórica, 612, 631
Correlación poliserial, 612, 613
Correlación semiparcial, 143, 182
Correlación tetracórica, 612, 631
Correlaciones estructurales: véase Cargas discriminantes
CORRESP, 576
COSAN, 634
Covariaciones teóricas: véase Análisis multivariante de la varianza
COVRATIO, 207, 213, 222
Criterio aleatorio en el análisis discriminante
  aleatoriedad máxima, 250, 272-273
  aleatoriedad proporcional, 273
  estadístico Q de Press, 270, 307
Criterio de aleatoriedad proporcional, 250
Criterio de información de Akaike (AIC), 685
Criterio de la varianza, porcentaje de: véase Porcentaje del criterio de la varianza
Criterio de máxima aleatoriedad: véase Criterio de aleatoriedad, máximo
Criterio de máxima verosimilitud: véase Lambda de Wilks
Criterio de Pilliai
  en el análisis de correlación canónica, 476
  en el análisis del criterio múltiple, 266
  en el análisis multivariante de la varianza, 366
Cuadrados menores, 143
Curtosis, 31, 64
Curva logística, 250, 282
CHAID, 718
Chi-cuadrado, 282-284, 572-574
Chi-cuadrado normada (SEM), 684-685
  En la escala multidimensional (MDS), 547

D

Datos
  definición, 713-714
  incremento de, 708
  preparación para la red neural, 724-725
Datos agregados, 707, 713-714
Datos analíticos, 707, 713
Datos censurados, 31, 41-42
Datos confundidos, 547
Datos cuantitativos: véase Datos medibles
Datos de preferencia (MDS), 547, 554, 559-560, 568-569
  análisis interno frente a externo, 567
  método de comparación por pares, 560
  método de ordenación directo, 560
  recogida de datos, 559-560
  representación puntual frente a vectorial, 568-569
Datos de semejanza (MDS), 547, 554, 558-559
  datos confundidos, 559
  diferencia desde preferencias, 560
  medidas derivadas, 559
  recogida de datos558-559
Datos dicotómicos: véase Variable binaria
Datos diferentes: véase Datos semejantes
Datos métricos, 2
Datos no métricos, 2-3
Datos operacionales, 707, 713
Definición conceptual, 79
Dendrograma, 491, 510, 524
Depóstito, 707
Descubrimiento de conocimientos de bases de datos (KDD), 707
Desigualdad de Bonferroni, 346, 369
DFBETA, 207, 213
DFFIT, 207, 222
Diagrama de árbol: véase Dendrograma
Diagrama de perfil, 491
Diagrama de tallo y hojas, 31, 33-34
Diagrama ículo vertical, 491
Diagramas de relaciones, 612, 615-616, 675-677
  conversión a la notación de LISREL, 674-677
  ejemplo, 653-668
  intercorrelación entre constructos, 622-623
  paso a ecuaciones estructurales, 624-628, 656
  representación de relaciones causales, 615-616
Diferencia honestamente significativa (HSD) de Tukey, 370
Diferencia menos significativa (LSD) de Tukey, 370
Dimensión inicial, 547
Dimensiones, 547, 548, 767, 717
  en el análisis de correspondencia, 577
  en el análisis de factores, 105
  en el análisis discriminante múltiple, 252-259
  en el análisis multidimensional, 547, 563-565, 569-571
Dimensiones percibidas, 547, 548
Dimesiones objetivas, 547
Diseño de experimentos, 346
Diseño de puente, 408, 432
Diseño equilibrado, 408, 430
Diseño factorial fraccional, 408, 430
Diseño factorial
  en el análisis conjunto, 408, 430
  en el análisis multivariante de la varianza, 346, 357-358, 388-395
Disparidades, 547, 564
Distancia D^2 de Mahalanobis
  en el análisis cluster, 491, 506
  en el análisis discriminante múltiple (MDA), 287-289
  en el análisis discriminante múltiple, 266, 267
  en la detección de atípicos, 59
  en la regresión múltiple, 207, 212
Distancia de Cook, 207, 213, 221
Distancia euclídea, 491, 495, 504
Distribución de término error normal: véase Error
Distribución normal multivariante, 347, 364
Distribución normal, 31
División de la muestra: véase Análisis discriminante múltiple

E

Efecto de tipo respuesta, 491, 507
Efecto interacción
  en el análisis conjunto, 408, 424-426
  en el análisis multivariante de la varianza, 346, 358-359, 360, 394
  en la regresión múltiple, 164-165
Efecto moderador, 143, 164-165
Efecto tamaño, 2, 8, 346, 367-368
Efectos principales
  en el análisis conjunto, 408
  en el análisis multivariante de la varianza, 346, 358
Elemento de datos, 707
Eliminación progresiva, 142-143, 171-172
Encuestas, administración de, 432-433
Enfoque conjunto basado en la elección, 408, 440-443
  ejemplo hipotético, 612-613
Enfoque de búsqueda secuencial, 171-173
Enfoque de construcción de modelos multivariantes, 22-24
Enfoque de disponibilidad completa (observaciones ausentes), 31, 44-45
Enfoque de distancia absoluta, 491, 504
Enfoque de los extremos polares, 250, 262
Enfoque de observaciones completas, 31, 44
EQS, 628, 634
Error
  control del porcentaje de error en la experimentación, 354
  error de especificación, 2, 21, 612, 621-622
  error de medida en el modelo de ecuaciones estructurales, 612, 613-614
  error estándar del coeficiente, 175
  error estándar del valor estimado, 149-150, 191
  errores de mensajes informáticos, 634
  estándar, 143, 346-348
  independencia de los términos de error, 195-196
  normalidad de la distribución del término de error, 168
  residuo en la regresión múltiple, 166-167
  término de error de varianza constante: véase Homocedasticidad
Error de aproximación cuadrático medio (RMSEA), 682

Error de especificación, 2, 21
en el análisis de regresión múltiple, 143, 158
en modelos de ecuaciones estructurales, 612, 621-622
Error de medida, 2, 7-8, 79, 104
en el modelo de ecuaciones estructurales, 612, 513-614
en la regresión múltiple, 143, 158
Error de predicción, 143
en la regresión múltiple, 143
Error de tipo I , 2, 7-10, 346, 348
Error de tipo II, 2, 8, 346, 367
Error estadístico, 7-10
Error estándar del coeficiente: véase Error
Error estándar del estimate: véase Error
Error estándar: véase Error
Escala de ratio, 6-7
Escala de semejanza, 547, 549
Escala métrica, 6-7
Escala no métrica, 5-6
Escala nominal, 6
Escalas de medida, 5-7
Escalas ordinales, 6
Escalas sumadas, 2, 7, 80, 104-106, 116-118
Espacio de búsqueda, 707
Estadístico de Wald, 250, 284
Estadístico F, 348
Estadístico Press Q
en el análisis discriminante múltiple, 250, 292
en la regresión multivariante, 143, 186
Estadístico R: véase Coeficiente de determinación
Estadístico ratio de verosimilitud chi-cuadrado (SEM), 680-681
Estadístico t, 346, 347
Estadísticos U: véase Lambda de Wilks
Estimación de máxima verosimilitud (MLE), 612, 632
Estimación de modelo
en la regresión logística, 318, 319
en las redes neurales, 725-726
Estimación del modelo, 724-727
Estimación directa, 633
Estimación por etapas
en el análisis discriminante múltiple, 250, 264, 287-288, 300-311
en la regresión logística, 319-321
en la regresión múltiple, 143, 171
Estimación simultánea
en el análisis discriminante múltiple, 250, 264
Estimaciones infractoras (SEM), 612, 637
Estímulo, 408
Estímulos de validación, 408
Estrategia de modelización corroborante, 613, 618-619
Estrategia de modelización en SEM, 612, 618-620
desarrollo del modelo, 632-634
modelización corroborante, 618-619
modelos rivales, 619
Estructura de preferencias, 408, 409-410
Evaluación subjetiva, 547, 564
Exposición de preferencias, 547, 567
Extracción de datos, 707, 709-711, 714-721
exploración frente a confirmación, 716

F

Factor de inflación de la varianza (VIF), 144, 183, 185, 208, 209
Factor en bloques, 346, 358
Factor
en ANOVA/MANOVA, 346
en el análisis conjunto, 408
en el análisis de factores, 79, 72-74
Fase de preparación de redes neurales, 720
Fiabilidad, 2, 7, 80, 105-106
en el modelo de ecuaciones estructurales, 612, 613, 639
Función canónica, 469, 474-476, 481-482
Función de activación, 707, 722
Función de base radial, 707, 722
Función de clasificación, 267
Función de distancia normalizada, 491, 507
Función discriminante lineal de Fisher, 267
Función discriminante, 350
en el análisis discriminante múltiple, 250
en el análisis multivariante de la varianza, 346
Función Sigmoid, 707, 722

G

Glyphs, 37
Grados de libertad
en el análisis multivariante de la varianza, 348
en el modelo de ecuaciones estructurales, 612-613, 635
en la regresión múltiple, 143, 174
Gráfico de dispersión, 31, 35, 36, 60
Gráfico de no-correlación, 143, 167
Gráfico de regresión parcial, 143, 166-167
Gráficos espaciales: véase Mapas perceptivos
Grupo de entropía, 491, 497

H

Heterocedasticidad
en la regresión múltiple, 167-168
Hipótesis nula, 346, 348
Histograma, 31, 33-34
HOMALS, 576
Homocedasticidad, 31, 66-68
contrastes estadísticos, 67
contrastes gráficos, 67
en la regresión múltiple, 102
ilustración de, 73-74
remedios para la heterocedasticidad, 67

I

Identificación, 612, 635-637
Igualdad de las matrices de covarianzas entre grupos 363-364
Independencia del término de error: véase Error
Independencia, 346
Indeterminación factorial, 79, 91-92
Indicador, 2, 7, 14, 80, 104, 612
Índice ajustado normado (NFI) (SEM), 683
Índice de ajuste, 547, 564
Índice de bondad de ajuste corregido (AGFI), 649, 683
Indice de condicionamiento, 207, 208

Índice de potencia, 250, 277-279, 312-313
Índice de redundancia (CCA), 469-470, 477-478
Índice de Tucker-Lewis (TLI) (SEM), 683
Índice GFI (SEM), 681
Índices de modificación, 612-613, 641
INDSCAL, 557, 585
Inferencia paramétrica, 729
Inteligencia artificial (AI), 707, 710
Intenciones, 256-259
Interacción disordinal, 346, 359
Interacción ordinal, 346, 359
Intervalos de confianza, 707, 718

J

Jitter, 707

K

KYST, 557

L

Lambda de Wilks
  en el análisis discriminante múltiple, 266
  en el análisis multivariante de la varianza, 347, 366
Lenguaje de petición estructurada (SQL), 707
Limpieza de datos, 707
Linealidad, 31, 67, 74-75
  análisis gráfico de los residuos, 167
  en el análisis multivariante de la varianza, 364
  en el modelo de ecuaciones estructurales, 624
  en la regresión múltiple, 143, 160-161, 167-168, 193-195
  identificación, 67
  ilustración de, 74-75
  polinomiales, 163-164
  remedios para la no linealidad, 67
LISREL, 14, 612, 628
  en el modelo de ecuaciones estructurales (SEM), 636
  notación, 671-673
  uso en el modelo de ecuaciones estructurales, 624, 634, 636, 669
LVPLS, 634

M

MANCOVA: véase Análisis multivariante de la varianza
Mapa perceptivo, 547
Mapa territorial, 275, 310-311
MAPWISE, 576
Matriz de asignación: véase Matriz de clasificación
Matriz de cargas: véase Cargas discriminantes
Matriz de clasificación
  en el análisis discriminante múltiple, 17, 268-274
  en la regresión logística, 255
Matriz de confusión: véase Matriz de clasificación
Matriz de correlación anti-imagen, 79, 88
Matriz de correlaciones
  en el análisis de factores, 79
  en los modelos de ecuaciones estructurales, 628
Matriz de descomposición de la varianza del coeficiente de regresión, 207, 208-209
Matriz de pauta factorial, 79, 101
Matriz de varianzas covarianzas
  igualdad entre grupos en el análisis discriminante, 264-265, 286
  igualdad entre grupos en MANOVA, 363-364
  uso como matriz de entrada en el modelo de ecuaciones estructurales, 628-632
Matriz estructural canónica, 485
Matriz factorial, 79, 95
  interpretación de, 100-102
Matriz sombrero, 267, 212
Mayor raíz característica (GCR)
  en el análisis de correlación canónica, 476
  en el análisis multivariante de la varianza, 346, 350
Mayor raíz característica de Roy: véase Mayor raíz característica
MDPREF, 556, 567
MDSCAL 557, 567
Medida de adecuación muestral, 79, 88-89
Medida de tensión, 547, 564
Medida extraída de la varianza, 612, 639
Medida multivariante, 2, 7
Medidas absolutas de ajuste, 612, 680-682
Medidas de ajuste incremental, 683
Medidas de ajuste parsimonioso, 684-685
Medidas de distancia (análisis de conglomerados)
  cálculo, 502-506
  consecuencia de las escalas de las variables, 506
  distancia de Mahalanobis, 506
  medida de distancia absoluta, 504
  medida de distancia euclídea, 495, 504
  uso como medidas de semejanza, 495, 502-506
Medidas de semejanza, 495, 501-507
  distancia absoluta, 504
  distancia de Mahalanobis, 506
  distancia euclídea, 495, 504
  medidas de asociación, 79-80
  medidas de correlación, 502, 503-504
  medidas de distancia, 502-506
Medidas derivadas, 547, 559
  en el análisis multidimensional, 547-548
Medidas repetidas346, 362
Metadatos, 707, 714
Método de comparación por pares, 408, 429
Método de composición, 408
  en el análisis conjunto, 408, 415
  en el análisis mltidimensional, 547, 557-558
Método de conector único, 491, 511
Método de descomposición
  en el análisis conjunto, 415
  en el análisis multidimensional, 547, 556-558
Método de perfil completo, 408, 427-429
Método de umbral paralelo, 491, 514
Método de umbral secuencial, 491, 513
Método de Ward, 492, 513, 522-523, 524
Método del centroide, 491, 513
Método del intercambio, 408, 427, 429
Método divisivo, 491, 510
Método U, 279

Métodos combinatoriales (regresión múltiple), 173
Métodos de aglomeración, 491, 495-596
Métodos de conector (análisis de de conglomerados), 511-514
  conector completo, 511-512
  conector medio, 513
  conector único, 511
  método de Ward, 513
  método del centroide, 513
Métodos de imputación para observaciones ausentes, 31, 44-54
  enfoque de casos completos, 44
  enfoque de disponibilidad completa, 44-45
  ilustración de, 48-54
  imputación *cold deck*, 46
  métodos de regresión, 46-56
  métodos múltiples, 46-47
  procedimientos basados en el modelo, 48
  sustitución de casos, 45-46
  sustitución por la media, 46
MINISSA, 557
Modelo, 612, 614, 617
Modelo adaptativo en el análisis conjunto, 408, 439-440
Modelo aditivo, 408, 424
Modelo de decomposición, 415
Modelo de descomposición, 408
Modelo de ecuaciones estructurales (SEM), 14, 611-690
  bondad de ajuste del modelo de medida, 640, 662
  bondad de ajuste del modelo estructural, 640, 662-665
  desde las diagramas de relaciones hasta la notación LISREL, 674-677
  diagramas de secuencias, 615-616
  ejemplo hipotético, 614-617
  ejemplo, análisis de factores corroborante, 642-653
  ejemplo, modelo estructural, 653-668
  estimaciones fiables de constructos, 627
  estrategia de modelización, 618-620
  función de la teoría, 614, 617-618
  identificación del modelo, 635-637
  índice de validación cruzada, 682
  interpretación, 641-642
  matriz de entrada, 628-632
  medidas absolutas de ajuste conjunto, 680-682
  medidas de bondad de ajuste conjunto, 637-638, 659-662, 677-690
  medidas de bondad de ajuste parsimonioso, 684-685
  medidas incrementales de bondad de ajuste, 683
  métodos de estimación, 631-632
  modelo de medida, 625-626
  modelo estructural, 624-625
  modelos rivales o anidados, 640-641, 664-666
  notación LISREL, 671-673
  notación matemática, 671-677
  proceso de decisión, 620-642
  relaciones causales, 620-622
  supuestos, 624
  tamaño muestral, 631
Modelo de independencia, 612
Modelo de medida (SEM), 612, 614, 617
  bondad de ajuste, 632, 640, 649-650
  ejemplo de ordenador, 642-653
  semejanza con el análisis de factores, 625-626
Modelo de probabilidad lineal, 13-14
  semejanza con la regresión múltiple, 284
  significación estadística de los coeficientes, 284
  uso de una variable dependiente binaria, 280-284
Modelo estructural (SEM), 612
  bondad de ajuste, 684
  conversión desde el diagrama de secuencias, 626, 674-677
  ejemplo, diagrama de secuencias, 653-668
  notación LISREL, 674-677
Modelo Kohonen, 707, 722
Modelo nulo 143, 167
Modelo recién identificado, 612, 635
Modelo sobreidentificado, 612, 635
Modelo subidentificado, 636
Modelos anidados, 612, 619, 641
Modelos conjuntos autoexplicados, 408
Modelos conjuntos híbridos, 439-440
Modelos rivales
  como estrategia de modelización, 612, 619
  uso de, 640-641, 664-666
Muestra de preparación, 707
Muestra de validación
  en el análisis de correlación canónica, 479-480
  en el análisis de regresión múltiple, 186
  en el análisis discriminante múltiple, 263
Multicolinelidad, 2, 21, 79, 88, 208-209
  cálculo de la varianza compartida y única, 182-183
  en el análisis cluster, 491, 508
  en el análisis de regresión múltiple, 143, 151-152, 181-185, 199-200, 207
  en el análisis discriminante múltiple, 264, 276
  en el análisis multivariante de la varianza, 364
  factor de inflación de la varianza (VIF), 199, 209
  ilustración de, 209
  influencia sobre la varianza compartida y única, 182-183
  singularidad, 184
  tolerancia, 209
  valoración con una matriz de descomposición de la varianza, 208-209
MULTISCALE, 557

N

Nivel de significación alfa, 2, 8-10, 346
Nivel de significación, 346
Nivel de variable independiente en el análisis conjunto, 408, 423-424
Nodo, 707, 719-720, 722
No-recursiva, 612
Normalidad de variables, 31, 63-66
  análisis gráfico, 64
  asimetría, 64
  contrastes estadísticos, 64-66
  curtosis, 64
  en el análisis factorial, 88
  en el análisis multivariante de la varianza, 364
  en la regresión múltiple, 196

en los modelos de ecuaciones estructurales, 624
ilustración, 70-73
tratamiento de la no normalidad, 65

O

Objeto, 547, 548
en el análisis cluster, 491
Observaciones ausentes prescindibles, 31, 41-42
Observaciones ausentes, 32, 39-44
diagnosis, 41-43
ilustración de, 48-54
MAR (ausencia aleatoria), 43
MCAR (ausencia totalmente aleatoria), 43
mecanismo de pérdida de datos, 31, 39-40, 43
métodos de imputación, 44-47
observaciones ausentes prescindibles frente a no prescindibles, 41-42
razones, 39-40
tratamiento, 44-57
Observaciones influyentes, 143, 177-178, 207, 612
comparación de residuos, 214-216
COVRATIO, 213
DFBETA, 213
distancia de Cook, 213
distancia Mahalanobis, 212
ejemplo de la base de datos HATCO, 219-220
identificación por medio de los residuos, 210-211
matriz sombrero, 150
medidas de diagnóstico, 219-220
puntos de apalancamiento, 211-212
SDFBETA, 221
SDFFIT, 213
Ortogonalidad
en el análisis conjunto, 408, 430
en el análisis de correlación canónica, 469, 476
en el análisis de factores, 79, 95
en el análisis multivariante de la varianza, 346, 351

P

Para la clasificación, 727-728
Parámetro, 143
Parrilla de importancia/realización, 547-557
Parsimoni
Aen las técnicas multivariantes, 21
en el modelo de ecuaciones estructurales, 612, 638
Perceptor multicapa, 707, 722
Perfiles multivariantes
representaciones gráficas, 37-38
uso en el análisis de los datos, 37
Petición, 707, 717
Polinomiales, 143, 163-164
POLYCON, 557
Ponderación discriminante, 250, 276
Ponderaciones canónicas, 478
Porcentaje correctamente clasificado, 255
Porcentaje del criterio de la varianza, 93
Potencia, 2, 7-10
componentes, 7-10
en el análisis de regresión múltiple, 143, 159
en el análisis multivariante de la varianza, 346, 367-368
influencia del tamaño muestral, 8-9, 368
relación con los errores de tipo I y de tipo II, 7-8, 367
Precio, 423
Precisión al clasificar: véase Matriz de clasificación
Preferencia, selección de una medida de, 432
PREFMAP, 557, 567
Procedimiento de optimización, 491, 514
Procedimientos de cluster jerárquicos, 491, 495, 510-514
Procedimientos de conglomerados no jerárquicos, 491, 513-514
método de umbral paralelo, 514
método de umbral secuencial, 513
técnicas de optimización, 514
uso de semillas cluster, 513
Procesamiento analítico on-line (OLAP), 707, 713, 717
Proceso de decisión en la construcción de modelos multivariantes
definición, 22-24
en el análisis conjunto, 416-418
en el análisis de conglomerados, 497-498
en el análisis de correlación canónica, 473-474
en el análisis de factores, 82-107
en el análisis discriminante múltiple, 260-280
en el análisis multidimensional, 552-571
en el análisis multivariante de la varianza, 354-372
en el modelo de ecuaciones estructurales, 628-634
en la regresión múltiple, 153-155
PROFIT, 587
Propagación progresiva, 707, 723
Proyecciones, 547
Pseudo R2, 250
Punto ideal, 547, 565-568
estimación interna frente a externa, 567
representación puntual frente a vectorial, 568-569, 583
Puntos de apalancamiento, 143, 179, 207, 211-212, 216
Puntuación de corte óptima, 250
Puntuación de corte, 250
cálculo, 291-293
uso en la clasificación, 252-269
Puntuación discriminante Z, 251, 276
Puntuación inversa, 80, 117-118
Puntuación Z, 250, 251
Puntuaciones factoriales, 79

R

R múltiple, 191
Raíces canónicas, 469, 476
Raíces latentes, 613
Raíz latente, 79, 92-93
Ratio de aciertos, 250, 273-274
Ratio F, 174
Ratio impar, 250, 282
Redes neurales, 710, 719-728
Regla de composición, 424-426
Regla de parada, 516
en el análisis cluster, 491
Reglas de asociación, 707, 718
Regresión logística, 250

como alternativa al análisis discriminante múltiple (MDA), 250, 316-322
ejemplo ilustrativo de, 316-322
Regresión múltiple, 11, 143-225
aplicaciones, 143-144
coeficientes beta, 180-181
coeficientes de regresión, 175-177, 180, 191
diagnosis avanzada, 207-225
diseño de la investigación, 158-166
efecto de interacción o efecto moderador, 164-165
efectos no lineales, 160-161
ejemplo muestral, 730-733
estadístico PRESS, 186
estimación por etapas, 192-193
estimación y ajuste de modelo, 169, 171-180
grados de libertad, 174
ilustración de, 187-204
interpretación, 180-185
modelo confirmatorio, 201-202
multicolinealidad, 151-152, 181-185, 199-200
objetivos, 155-158
observaciones influyentes, 196-198, 207
potencia, 159
predicción frente a explicación, 155-157
proceso de decisión, 153-155
requisitos de tamaño muestral, 158-160
selección de variables dependientes e independientes, 157-159
selección de variables, 169, 171-173
significación estadística, 174-177
supuestos, 166-169, 173-174, 188-189, 193-196
transformaciones, 161
uso de una única variable independiente, 146-151
uso de varias variables independientes, 151-153
validación del modelo, 186-187
validación, 115-116, 143-144
variables ficticias para datos no medibles, 161-162
varianza única y compartida, 182-183
Regresión simple, 143, 145, 146-151
Regresiones parciales combinando variables, 143, 173
Relación causal, 612
Relación estadística, 143, 157
Remuestreo, 707, 710, 728-733
ejemplo de, 730-733
métodos, 729-730
Replicación, 346, 371-272
Representación gráfica multivariante, 31, 37-38
Representación vectorial (MDS), 568-569, 586
Representaciones puntuales, 568-569
Residuo cuadrático medio (RMSR) (SEM), 682
Residuo tipificado, 207, 211
Residuos basados t de Student, 144, 166-167, 208, 211
Residuos, 31, 35, 207, 210-211
comparación de, 214-216
en el análisis de regresión múltiple, 143, 166
Rotación de factor oblícua, 79, 95
Rotación de resultados
en el análisis de factores, 95-99, 112-113
en el análisis discriminante múltiple, 277, 311-312
EQUIMAX, 98
QUARTIMAX, 97
rotación oblícua, 95, 114-115
rotación ortogonal, 95, 96, 97-98
VARIMAX, 97-98, 112-113
Rotación EQUIMAX, 98
Rotación factorial, 79, 95-99
Rotación VARIMAX, 80

S

SDFFIT, 213
Semejanza inter-objetos, 491
Semillas de cluster, 491, 513
Sesgo de nodo, 707, 725
Significación estadística, 7-10, 20
en la regresión logística, 317
Significación práctica, 2, 19-20
Simetría, 31, 64
Simulación, 633-634
en el análisis de modelos de ecuación estructural, 612
Simulador conjunto: véase Simulador aleatorio
Simulador de elecciones, 408, 449-450
ejemplo de ordenador, 449-450
estimación, 437-438
Singularidad, 143
Sistema de apoyo a la decisión (DSS), 707, 713
Sistema operacional, 707, 713
Sistemas de apoyo a la decisión (DSS), 711
Solución degenerada, 547, 583
en el análisis multidimensional, 547
Suma del cuadrado de la regresión (SSR), 144, 150
Suma del cuadrado de los errores (SSE), 144, 146
Suma total de los cuadrados (TSS), 144, 150
Supervisados, 707
Supuestos en el análisis multivariante, 63-75
ausencia de correlación en los errores, 67-68
homocedasticidad, 66-67, 167
independencia de las observaciones, 363
linealidad, 67, 167-168
normalidad, 63-66, 168
variables individuales frente a valor teórico, 63, 166-167

T

T^2 de Hotelling, 346, 349-350, 375-380
Tabla de contingencia, 547, 571
Tabla de tabulación cruzada, 547
Tamaño muestral, 8-9, 20
en el análisis discriminante múltiple, 263
en el análisis factorial, 87-88
en el análisis multivariante de la varianza, 357
en el modelo de ecuaciones estructurales, 631
en la regresión múltiple, 158-160
en las redes neurales, 724
Taxonomía numérica: véase Análisis cluster
Taxonomía, 491, 498
Técnica de dependencia, 2, 15
Técnica de interdependencia, 2, 15
Técnicas estadísticas multivariantes
clasificación, 15-19
definición 2

distinción entre técnicas de dependencia y de interdependencia, 15-18
selección, 16-17
Teoría, 612, 614, 817-618
Término de error de varianza constante: véase Homocedasticidad
TETRAD, 619
Tipificación centrada por filas, 491-507
Tipificación intrasujeto, 492, 507
Tipificación, 143, 507-508
Tipología: véase Análisis cluster
Tipos de, 722
Toleranci
Aen el análisis discriminante múltiple, 250
en la regresión múltiple, 144, 183, 185, 208, 209
Transacción, 708
Transformación de datos, 31-32, 68-70
construcción de variables ficticias, 161-162
efecto de interacción y efecto moderador, 164-165
orientaciones, 70
para conseguir homocedasticidad, 68, 69
para conseguir linealidad, 69-70
para conseguir normalidad, 68-69
Transformación logit, 250
Transformación, 144
Tratamiento, 2, 12
en el análisis conjunto, 408
en el análisis multivariante de la varianza, 346
Traza de Hotelling
en el análisis de correlación canónica, 476
en el análisis discriminante múltiple, 266
en el análisis multivariante de la varianza, 366
Traza de probabilidad normal, 31, 63-66, 143, 168
Traza, 80

U

Unidimensionalidad, 612, 638
Utilidad, 408

V

Validación con división de la muestr
Aen el análisis discriminante múltiple, 250, 279-280
en la regresión múltiple, 200-201
Validación cruzada, 250, 682
Validación, 2, 7, 80, 106-107, 612, 639
Validez de contenido, 79
Validez predictiva, 518
Valor crítico, 346, 348
Valor de partida, 636
Valor de un atributo, 408, 410-414, 447-448
cálculo, 447-448
evaluación, 427-448
Valor de verosimilitud, 250, 282, 283
Valor extremo: véase Atípico
Valor teórico conjunto, 408
Valor teórico de la regresión, 143, 145
Valor teórico
en el análisis conjunto, 408, 415
en el análisis de cluster, 491, 492, 498-500
en el análisis de correlación canónica, 469
en el análisis de factores, 81
en el análisis discriminante múltiple, 250-251
en el análisis multivariante de la varianza, 346, 348-349
en la regresión múltiple, 145, 166-167
Valores F (o t) parciales
en el análisis de regresión multivariante, 143
Valores teóricos canónicos, 469, 472, 482-485
Variable
latente, 612
manifiesta, 612
métrica, 250
Variable binaria
como variable dependiente en la regresión logística, 280-284
como variable ficticia, 75-77
Variable categórica, 250
uso en la regresión múltiple, 250-251
Variable cualitativa: véase Datos no medibles
Variable de criterios: véase Variable dependiente
Variable dependiente, 2, 15
en el análisis conjunto, 430-432
en el análisis discriminante múltiple, 262
en el análisis multivariante de la varianza, 348-349, 354
en la regresión múltiple, 75-77, 143, 144
Variable ficticia, 2, 18, 31, 75-77, 79, 161-162
codificación de efectos, 162
codificación de indicadores, 162
en la regresión múltiple, 143
uso, 161-162
Variable independiente, 2, 15
en el análisis de correlación canónica, 415
en el análisis de regresión, 143, 144
en el análisis discriminante múltiple, 250, 262
en el análisis multivariante de la varianza, 348-349
Variable manifiesta, 612, 613
Variable predictor aleatorio (regresión múltiple), 160
Variable predictor: véase Variable independiente
Variable suplente, 80, 116
Variable/constructo no observable: véase Variable/concepto latente
Variables de interacción: véase Efecto interacción
variables de predictor fijo (regresión múltiple), 160
Varianza compartida, 79, 91
Varianza común, 79, 89
Varianza del error en el análisis de factores, 79
Varianza específica, 80, 84
Varianza intragrupos, 348
Varianza única
en el análisis factorial, 89
influencia de la multicolinealidad, 151-152
Vecino más alejado: véase Conector completo
Vecino más cercano: véase Método de vínculo único
Vector
en el análisis discriminante múltiple, 250, 278
en el análisis multidimensional, 547, 568-569
en el análisis multivariante de la varianza, 386
Visualización de datos, 707, 717