Tema 10

Análisis factorial

El análisis factorial es una técnica cuyo propósito es definir la estructura subyacente de un conjunto de datos, de tal manera que analiza la estructura de correlación entre las variables mediante la definición de una serie de dimensiones subyacentes llamadas *Factores*. Esto se realiza con un doble fin: el resumen y la reducción de los datos. A la hora de resumir se obtiene unas dimensiones que cuando son interpretadas describen los datos de una manera más sencilla; con respecto a la reducción, se pueden obtener los cálculos de las puntuaciones para cada dimensión y sustituirlos por las variables originales.

La ventaja que tiene el Análisis Factorial es que se tratan todas las variables simultáneamente, cada una relacionada con las demás y no existe una variable dependiente y otras independientes.

10.1. El análisis factorial

El objetivo fundamental será el encontrar una manera de resumir la información contenida en una serie de variables originales, mediante una serie de dimensiones compuestas (factores) o valores teóricos con una mínima perdida de información.

10.1.1. Hipótesis previas

Dentro de las hipótesis previas la primera de ellas sería la normalidad, aunque esta solo será necesaria si se realizan pruebas estadísticas de significación. También es deseable que exista un cierto grado de multicolinealidad, ya que el objetivo es identificar variables relacionadas.

El análisis factorial debe tener suficientes correlaciones para poder aplicarse. Si no hay una número sustancial de correlaciones mayores de 0.3 entonces es probablemente inadecuado. Las correlaciones entre variables pueden ser analizadas mediante el cálculo de las correlaciones parciales de tal manera que si las correlaciones parciales son bajas, entonces no existen factores subyacentes verdaderos y el análisis factorial es inapropiado. Esta hipótesis también se puede contrastar buscando que los valores de la diagonal de la matriz anti-imagen sean altos.

10.1.2. Estimación de los factores

Para decidir el número de factores que se deben extraer, utilizaremos el porcentaje de la varianza total de los datos, de tal manera que el porcentaje de varianza debe explicar un valor que consideremos válido. Por ejemplo, en ciencias naturales se toman factores hasta explicar al menos un 95 % de la varianza, frente a ciencias sociales que es normal considerar sobre el 60 %-65 % de la varianza total.

Las comunalidades serán el porcentaje de varianza de cada variable explicada por el número de factores escogido.

10.1.3. Interpretación de los factores

Para determinar la solución utilizaremos las cargas factoriales, que son el medio para interpretar la función que cada variable desempeña para definir cada factor. Son las correlaciones entre cada variable y el factor, de tal manera que indican la correspondencia entre cada variable y el factor. Las cargas con valores entre ±0,30 se consideran de nivel mínimo, mayores de ±0,40 son más importantes y de ±0,50 son significativas. Por lo tanto cuanto mayor sea el valor absoluto de la carga más importante es esa variable para interpretar el factor, por ejemplo, una carga de 0.30 implica una explicación del 10 % de la varianza del factor, y uno de 0,50 una explicación del 25 %.

10.1.4. Rotación de los factores

La rotación consiste en girar los ejes factoriales. La solución inicial extrae los factores según su importancia, de tal manea que el primer factor explica la mayor parte de la varianza (y por lo tanto estará en función de más variables) y los siguientes factores van explicando progresivamente menor porcentaje de varianza. Con la rotación se distribuye la varianza en otros factores para lograr un patrón de factores más simple y más significativo.

Existen muchos métodos de rotación y la selección de uno u otro dependerá de las soluciones, de tal manera que esta se simplifique el máximo posible.

10.2. Aplicación mediante SPSS

Como ejemplo de Análisis Factorial vamos a utilizar el fichero **Factorial**, el cuál contiene los valores de 17 indicadores de tipo sanitario para 23 países. El objetivo es realizar una análisis factorial para encontrar una estructura subyacente en estos datos.

Para realizar una análisis factorial tenemos que elegir los menús:

Analizar
Reducción de datos
Análisis factorial

obteniendo el siguiente menú:

Figura 10.1: Menú análisis factorial

10.2.1. Botón Descriptivos

Con este botón podremos obtener:

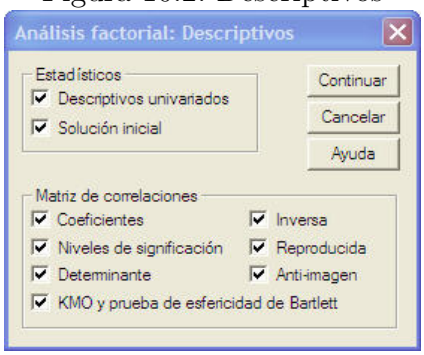
- *Estadísticos*:

  1. Descriptivos univariados: la media, la desviación típica y el número de casos válidos para cada variable.
  2. Solución inicial muestra las comunalidades iniciales, los autovalores y el porcentaje de varianza explicada.

- *Matriz de correlaciones*:

  1. Coeficientes, niveles de significación y determinante.
  2. Inversa, reproducida, anti-imagen
  3. KMO (medida de Kaiser-Meyer-Olkin) y prueba de esfericidad de Bartlett.

Figura 10.2: Descriptivos



Si seleccionamos todas las opciones, obtendremos los siguientes resultados:

- Con la opción estadísticos descriptivos obtenemos para todas las variables la media, la desviación típica, el número de casos válidos y si hay algún datos faltante.

Figura 10.3: Estadísticos

Estadísticos descriptivos

	Media	Desviación típica[a]	N del análisis[a]	N perdida
tasa de mortalidad cancer de pulmon	65,8217	24,40285	23	0
tasa de mortalidad por cirrosis hepatica	19,9652	14,75178	23	0
consumo de alcohol	9,4043	3,64623	23	0
heridos tráfico	4726,2609	2359,75105	23	0
tasa demortalidad perinatal	1,1800	,40925	23	0
tasa de mortalidad infantil	,9917	,29742	23	0
esperanza hombres	77,4696	1,57637	23	0
esperanza mujeres	70,7609	1,77117	23	0
dias cama	15,3522	4,08221	23	0
porcentaje de poblacion protegida ambulatorio	90,6957	20,06679	23	0
porcentaje de poblacion protegida hospital	95,7391	12,72155	23	0
poblacion por cama hospitalaria	113,6957	37,50204	23	0
gasto per cápita en médicos	739,6522	280,37793	23	0
pib en salud	7,3087	1,39281	23	0
pib gasto publico salud	5,7391	1,27698	23	0
tasa de admision hospitalaria	15,3087	4,07653	23	0

a. Para cada variable, los valores perdidos se sustituyen por la media de la variable.

- Con la matriz de correlaciones, su significación y el determinante. Es conveniente, como ocurre en este caso que los coeficientes de la matriz de correlaciones sean significativos en su mayor parte. También es interesante que el determinate de la matriz de correlaciones sea bajo, $1{,}35 * 10^{-10}$ en este ejemplo.

Figura 10.4: Matriz de correlaciones

Matriz de correlaciones[a]

		v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
Correlación	v1	1,000	,1	,319	,311	-,2	,009	-,357	,248	,182	,019	,055	,242	,183	,034	,016	,141
	v2	,064	1,0	,736	,315	,54	,535	-,330	,454	,213	,190	,000	,225	,211	,166	,005	,201
	v3	,319	,736	1,0	,347	,42	,440	-,456	,585	,167	,020	,116	,131	,161	,045	,046	,096
	v4	,311	,315	,347	1,000	,0	,190	-,290	,330	,114	,064	,300	,047	,122	,287	,258	,098
	v5	,181	,536	,420	,022	1,0	,886	-,495	,417	,443	,100	,066	,704	,437	,595	,401	,496
	v6	,009	,535	,440	,190	,89	1,0	-,664	,577	,241	,028	,058	,559	,241	,541	,396	,446
	v7	,357	,3	,5	,290	,5	,66	1,000	,852	,087	,252	,055	,169	,009	,412	,312	,099
	v8	,248	,5	,6	,330	,4	,58	,852	1,000	,203	,156	,079	,110	,219	,204	,093	,067
	v9	,182	,2	,2	,114	,4	,24	,087	,203	1,000	,106	,023	,428	,998	,324	,237	,338
	v10	,019	,190	,020	,064	,1	,03	,252	,156	,106	1,000	,734	,182	,113	,284	,456	,031
	v11	,055	,000	,1	,300	,1	,06	,055	,079	,023	,734	1,000	,366	,016	,455	,543	,230
	v12	,242	,225	,131	,047	,70	,559	,169	,110	,428	,182	,366	1,00	,432	,302	,231	,550
	v13	,183	,2	,2	,122	,4	,24	,089	,219	,998	,113	,016	,432	1,000	,319	,241	,331
	v14	,034	,2	,0	,287	,6	,54	,412	,204	,324	,284	,455	,302	,319	1,000	,916	,440
	v15	,016	,0	,046	,258	,4	,40	,312	,093	,237	,456	,543	,231	,241	,916	1,000	,443
	v16	,141	,2	,1	,098	,5	,45	,099	,067	,338	,031	,230	,550	,331	,440	,443	1,000
Sig. (unilateral)	v1		,366	,069	,075	,20	,483	,047	,129	,245	,466	,401	,133	,229	,436	,471	,260
	v2	,366		,000	,072	,00	,004	,058	,015	,165	,189	,489	,151	,166	,224	,491	,179
	v3	,069	,000		,052	,02	,018	,014	,002	,224	,464	,298	,276	,232	,419	,418	,331
	v4	,075	,072	,052		,46	,192	,089	,062	,302	,386	,082	,416	,288	,085	,117	,329
	v5	,204	,004	,023	,460		,000	,008	,024	,017	,325	,379	,000	,019	,001	,029	,008
	v6	,483	,004	,018	,192	,00		,000	,002	,134	,449	,396	,003	,134	,004	,031	,016
	v7	,047	,058	,014	,089	,01	,000		,000	,346	,123	,402	,220	,343	,028	,074	,327
	v8	,129	,015	,002	,062	,02	,002	,000		,176	,239	,360	,296	,158	,175	,336	,381
	v9	,245	,165	,224	,302	,02	,134	,346	,176		,315	,459	,021	,000	,066	,138	,057
	v10	,466	,189	,464	,386	,33	,449	,123	,239	,315		,000	,202	,304	,095	,014	,444
	v11	,401	,489	,298	,082	,38	,396	,402	,360	,459	,000		,043	,467	,015	,004	,145
	v12	,133	,151	,276	,416	,00	,003	,220	,296	,021	,202	,043		,020	,081	,144	,003
	v13	,229	,166	,232	,289	,02	,134	,343	,158	,000	,304	,467	,020		,069	,134	,061
	v14	,436	,224	,419	,085	,00	,004	,025	,175	,066	,095	,015	,081	,069		,000	,018
	v15	,471	,491	,418	,117	,03	,031	,074	,336	,138	,014	,004	,144	,134	,000		,017
	v16	,260	,179	,331	,329	,01	,016	,327	,381	,057	,444	,145	,003	,061	,018	,017	

a. Determinante = 1,35E-010

- KMO y prueba de esfericidad de Bartlett. Con la prueba KMO contrastamos que las correlaciones parciales sean pequeñas. Si es estadístico KMO es grande el procedimiento es adecuado, si es pequeño será inadecuado, este estadístico oscila entre 0 y 1.

  La prueba de Bartlett contrasta si la matriz de correlaciones es una matriz identidad como hipótesis nula, es decir, que las variables están incorreladas; en caso de aceptarse esta hipótesis indicaría que el modelo factorial es inadecuado.

- Matriz anti-imagen. En un buen modelo factorial la mayoría de los elementos no diagonales deben ser pequeños y los de la diagonal grandes. En nuestro ejemplo ocurre esta circunstancia por los que el modelo factorial será bueno.

Figura 10.5: KMO y Bartlett

KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,480
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	359,787
	gl	120
	Sig.	,000

Las pruebas de KMO y Bartlett, y el estudio de la matriz anti-imagen y la matriz de correlaciones y su significación resultan satisfactorias por lo que al análisis factorial puede proporcionarnos buenos resultados.

10.2.2. Extracción

Con este botón obtendremos el análisis factorial propiamente dicho. Las opciones más importantes son:

- *Método*: con esta opción seleccionaremos el método para extraer los factores. Las opciones disponibles son:

  1. Componentes principales.
  2. Mínimos cuadrados no ponderados.
  3. Mínimos cuadrados generalizados.
  4. Máxima verosimilitud.
  5. Factorización de Ejes principales.
  6. Factorización Alfa.
  7. Factorización Imagen.

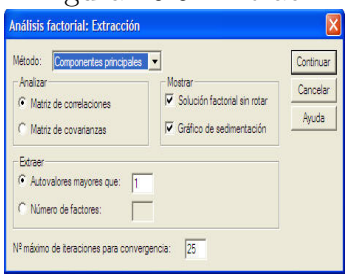
  Nosotros usaremos el método de componentes principales que consiste en realizar una combinación lineal de las variables, de tal manera que el primer componente principal sea la combinación que más varianza explique, el segundo la segunda mayor y que está incorrelado con el primero y así sucesivamente.

- *Analizar*. Permite especificar o una matriz de correlaciones o una matriz de covarianzas. Utilizaremos la matriz de correlaciones si las variables están en diferente escala y la de covarianzas en caso contrario.

- *Extraer*. Podemos elegir entre utilizar tantos factores como autovalores mayores de un valor. Se conoce como regla de Kaiser a quedarse con tantos factores como autovalores mayores que uno existan. También podemos indicar un número concreto de factores a extraer.

- *Mostrar*: ofrece la solución sin rotar y el gráfico de sedimentación.

Figura 10.6: Extraer



En este ejemplo, por lo tanto usaremos la *matriz de correlaciones*, el método de *componentes principales* y usaremos la ya comentada *regla de Kaiser*, es decir, nos quedaremos con tantos factores como autovalores mayores que 1. Los resultados son:

1. Las comunalidades son la proporción de varianza explicada por los factores de cada una de las variables. Vemos que todas las variables van a ser bien explicadas por la estructura factorial, salvo quizás las variables "PIB Gasto Publico en Salud" y "Heridos de Trafico".

Figura 10.7: Comunalidades

Comunalidades

	Inicial	Extracción
tasa de mortalidad cancer de pulmon	1,000	,816
tasa de mortalidad por cirrosis hepatica	1,000	,893
consumo de alcohol	1,000	,790
heridos tráfico	1,000	,649
tasa demortalidad perinatal	1,000	,920
tasa de mortalidad infanti	1,000	,853
esperanza hombres	1,000	,893
esperanza mujeres	1,000	,855
dias cama	1,000	,915
porcentaje de poblacion protegida ambulatorio	1,000	,808
porcentaje de poblacion protegida hospital	1,000	,904
poblacion por cama hospitalaria	1,000	,757
tasa de admision hospitalaria	1,000	,921
gasto per cápita en médicos	1,000	,935
pib en salud	1,000	,946
pib gasto publico salud	1,000	,531

Método de extracción: Análisis de Componentes princi

2. Con la tabla de Varianza Explicada obtenemos una tabla con todos los autovalores y el porcentaje de varianza explicada por cada uno de ellos. En principio tendremos tantos autovalores como variables tengamos. Mediante la regla de Kaiser tomamos tantos factores como autovalores mayores de 1, en este caso tomaríamos 5 factores. Estos cinco factores explican un 83.66 % de la varianza total. A partir de ahora surge la decisión del investigador, de tal manera que si este 83.66 % nos parece suficiente nos quedaríamos con esos cinco factores, si nos parece poco aumentaríamos algún factor y si queremos un modelo más sencillo eliminaríamos algún factor. Como vimos anteriormente, en problemas de tipo socioeconómico es suficiente con una explicación del 60-70 %, por lo que sería interesante para simplificar el problema quedarnos con una estructura de 3 o 4 factores; para ello en el botón extraer (10.2.2), indicaríamos el número concreto que nos interese.

Figura 10.8: Varianza explicada

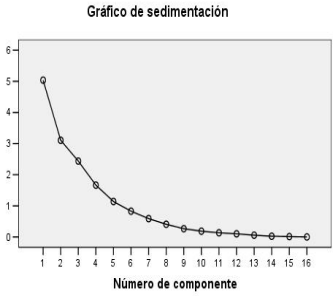
Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5,036	31,478	31,478	5,036	31,478	31,478
2	3,105	19,406	50,884	3,105	19,406	50,884
3	2,438	15,238	66,122	2,438	15,238	66,122
4	1,667	10,417	76,539	1,667	10,417	76,539
5	1,140	7,125	83,664	1,140	7,125	83,664
6	,829	5,184	88,848			
7	,590	3,685	92,532			
8	,409	2,557	95,089			
9	,267	1,671	96,760			
10	,187	1,166	97,926			
11	,135	,846	98,773			
12	,102	,635	99,408			
13	,057	,357	99,765			
14	,024	,151	99,916			
15	,013	,079	99,995			
16	,001	,005	100,000			

Método de extracción: Análisis de Componentes principales.

3. El gráfico de sedimentación es la representación gráfica de los autovalores, de tal manera que se suele usar para decidir el número de factores a utilizar, tomando factores hasta que la pendiente de la recta sea paralela al Eje X.

Figura 10.9: Gráfico de sedimentación



Por lo tanto, a partir del gráfico de sedimentación y principalmente del estudio de la varianza, nos quedaremos con una estructura factorial con 3 o 4 factores. Dependiendo de esta decisión las comunalidades serán:

Figura 10.10: Comunalidades con tres y cuatro factores

Comunalidades

	Inicial	Extracción
tasa de mortalidad cancer de pulmon	1,000	,299
tasa de mortalidad por cirrosis hepatica	1,000	,460
consumo de alcohol	1,000	,562
heridos tráfico	1,000	,364
tasa demortalidad perinatal	1,000	,850
tasa de mortalidad infantil	1,000	,819
esperanza hombres	1,000	,805
esperanza mujeres	1,000	,840
dias cama	1,000	,696
porcentaje de poblacion protegida ambulatorio	1,000	,507
porcentaje de poblacion protegida hospital	1,000	,814
poblacion por cama hospitalaria	1,000	,870
tasa de admision hospitalaria	1,000	,701
gasto per cápita en médicos	1,000	,867
pib en salud	1,000	,855
pib gasto publico salud	1,000	,471

Método de extracción: Análisis de Componentes principales.

Comunalidades

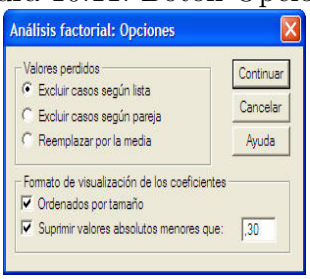
	Inicial	Extracción
tasa de mortalidad cancer de pulmon	1,000	,339
tasa de mortalidad por cirrosis hepatica	1,000	,686
consumo de alcohol	1,000	,772
heridos tráfico	1,000	,564
tasa demortalidad perinatal	1,000	,882
tasa de mortalidad infantil	1,000	,844
esperanza hombres	1,000	,854
esperanza mujeres	1,000	,850
dias cama	1,000	,862
porcentaje de poblacion protegida ambulatorio	1,000	,790
porcentaje de poblacion protegida hospital	1,000	,888
poblacion por cama hospitalaria	1,000	,754
tasa de admision hospitalaria	1,000	,869
gasto per cápita en médicos	1,000	,914
pib en salud	1,000	,885
pib gasto publico salud	1,000	,514

Método de extracción: Análisis de Componentes principales.

Evidentemente, con cuatro factores las comunalidades son más altas, el problema está en que este modelo es mas complejo que con tres factores, por lo que habrá que decidir entre estas dos estructuras, en nuestro ejemplo vamos a optar por una estructura con tres factores.

4. Una vez decidido el número de factores obtendremos la solución final que es la matriz de componentes, pero con una pequeña restricción. La matriz de componentes (también conocida como Factor Scores) son las cargas de cada variable en cada uno de los factores originados, de modo que las variables con cargas más altas en un factor, nos indica una estrecha relación entre la variable y ese factor, en definitiva es la correlación entre el factor y la variable. Por ello, solo nos interesarán aquellos valores que tengan altas cargas factoriales, para obtener estos valores, seleccionaremos el botón de *opciones* y seleccionaremos las pestañas de ordenar valores y suprimir aquellas con un valor menor de 0.30,

Figura 10.11: Botón Opciones



Resultando que el primer factor viene determinado por las variables Tasa de Mortalidad Perinatal, Tasa de Mortalidad Infantil, Gasto per Capita en Médico, Población por Cama, Esperanza de Vida Masculina, Tasa Hepática, PIB en Gasto Público, Consumo de Alcohol, Esperanza de Vida Femenina y PIB en Salud; el segundo factor viene determinado por Gasto per Capita en Médicos, Esperanza de Vida Masculina, Consumo de Alcohol, Esperanza de Vida Femenina, PIB en salud, Heridos de Tráfico, Porcentaje Protegido por Ambulatorio, Tasa de Cancer de Pulmón, Porcentaje Protegido por Hospital, Tasa de admisión Hospitalaria y Dias en Cama y el tercer factor por Gasto per Capita en médicos, Población por Cama Hospitalaria, Esperanza de Vida Masculina, Esperanza de Vida Femenina, PIB en salud, Porcentaje Protegido por Ambulatorio, Tasa de Cancer de Pulmón, Porcentaje Protegido por Hospital,Tasa de admisión Hospitalaria y Dias en Cama.

El problema está en que para la interpretación sería interesante que una variable tuviera una importante carga factorial para un solo factor y no dependiese de varios, para intentar que se de esta situación se recurre a las rotaciones factoriales.

Figura 10.12: Cargas factoriales

Matriz de componentes[a]

	Componente		
	1	2	3
tasa demortalidad perinatal	-,913		
tasa de mortalidad infantil	-,889		
gasto per cápita en médicos	,679	,498	-,398
poblacion por cama hospitalaria	-,646		-,491
esperanza hombres	,626	-,510	-,391
tasa de mortalidad por cirrosis hepatica	-,616		
pib gasto publico salud	,589		
consumo de alcohol	-,547	,506	
esperanza mujeres	,539	-,537	-,379
pib en salud	,540	,578	-,480
heridos tráfico		,545	
porcentaje de poblacion protegida ambulatorio		-,519	,485
tasa de mortalidad cancer de pulmon		,430	,338
porcentaje de poblacion protegida hospital		-,542	,721
tasa de admision hospitalaria	,497	,444	,506
dias cama	,501	,438	,504

Método de extracción: Análisis de componentes principales.
a. 3 componentes extraídos

10.2.3. Rotación

El fin de la rotación es ayudar a interpretar en los casos en los que las cargas factoriales no estén claras; las rotaciones simplemente son movimientos de los ejes de tal manera que se mantenga la comunalidad y el porcentaje de varianza explicada por el modelo, aunque no por cada eje. Lo que buscaremos será que cada factor tenga unos pocos pesos altos y el resto cercanos a cero, y que cada variable este saturada en solo un factor.

Para realizar una rotación elegiremos el botón *rotación*, el cuál nos permite escoger entre:

- *Método*: para elegir el método de rotación entre: varimax, equamax, quartimax, oblimin directo y promax.
- *Mostrar*:
  1. Solución rotada: se selecciona un método de rotación para obtener la solución rotada.
  2. Gráficos de saturaciones: muestra las representaciones tridimensional de las saturaciones factoriales para los tres primeros factores. Para una solución de dos factores, se representa un diagrama bidimensional. No se muestra el gráfico si sólo se extrae un factor.

En este caso, vamos a elegir una de las rotaciones que SPSS nos ofrece y concretamente la VARIMAX. Además marcaremos la solución rotada y el gráfico de saturaciones. Si vemos la tabla de la varianza explicada, observamos que no han cambiado la varianza total explicada (66,12 %) aunque si lo que explica cada factor, 31,478 % en el primer factor y sin rotar frente a 25,052 % rotado; 19,406 frente a 23,134 del segundo factor y 15,238 frente a 17,935 del tercer factor.

Figura 10.13: Varianza explicada

Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	5.035	31.470	31.470	5.035	31.470	31.470	4.008	25.052	25.052
2	3.105	19.406	50.864	3.105	19.406	50.864	3.702	23.134	48.187
3	2.438	15.238	66.102	2.438	15.238	66.102	2.670	17.915	66.102
4	1.667	10.417	76.519						

También vemos que como las soluciones rotadas se han simplificado:

Figura 10.14: Cargas factoriales rotadas

Matriz de componentes rotados[a]

	Componente		
	1	2	3
esperanza mujeres	-,904		
esperanza hombres	-,892		
tasa de mortalidad infantil	,753	-,477	
consumo de alcohol	,729		
tasa de mortalidad por cirrosis hepatica	,624		
tasa de admision hospitalaria		,831	
dias cama		,829	
poblacion por cama hospitalaria		-,775	
tasa demortalidad perinatal	,583	-,698	
pib gasto publico salud		,653	
tasa de mortalidad cancer de pulmon	,375	,395	
porcentaje de poblacion protegida hospital			-,869
pib en salud		,312	,849
gasto per cápita en médicos	-,312	,419	,770
porcentaje de poblacion protegida ambulatorio			-,701
heridos tráfico	,408		,444

En este caso el primer factor viene determinado por las variables Esperanza de Vida Masculina, Esperanza de Vida Femenina, Tasa de Mortalidad Infantil, Consumo de Alcohol, Tasa Hepática, Tasa de Mortalidad Perinatal, Tasa de Cancer de Pulmón, Gasto per Capita en Médico y Heridos de Tráfico; el segundo factor vendrá marcado por la Tasa de Mortalidad Infantil, Tasa de admisión Hospitalaria, Dias en Cama, Población por Cama, Tasa de Mortalidad Perinatal, PIB en Salud y Gastos per Capita en Médicos y el tercer factor determinado por Porcentaje Protegido por Hospital, PIB en Salud, Gasto per Capita en Médicos, Porcentaje Protegido por Ambulatorio y Heridos de Tráfico. En este caso los factores vienen determinados por menos variables y son más sencillos.

Por medio de los gráficos podemos deducir el sentido y significado de los factores. A continuación vemos los gráficos bidimensionales y tridimensionales. Estos gráficos son las coordenadas de cada variable con respecto a los factores, que como hemos visto son valores entre -1 y 1, siendo mayor la contribución cuanto mayor sea este valor. A partir de la posición de las variables con respecto a los ejes es donde se puede deducir el sentido y significado de los factores. Lo ideal es que no estén centrados en torno al punto (0,0) sino próximos al borde del gráfico, ya que la idea es identificar las variables con mayor peso factorial. Por ejemplo, con respecto a las componentes 1 y 2, vemos que las variables Esperanza de Vida Masculina y Femenina están muy juntas, por lo que están muy correladas entre si y a su vez con el factor 1 negativamente, y Consumo de Alcohol y Cirrosis Hepática lo están entre si y con el factor uno positivamente, por lo que podríamos concluir que el primer factor contrapone la Esperanza de

Vida con el Consumo de Alcohol.

Figura 10.15: Gráficos

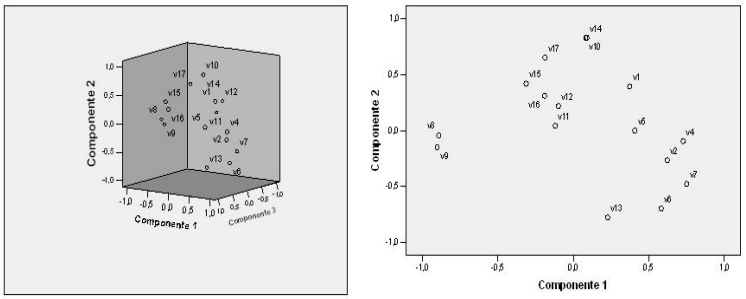
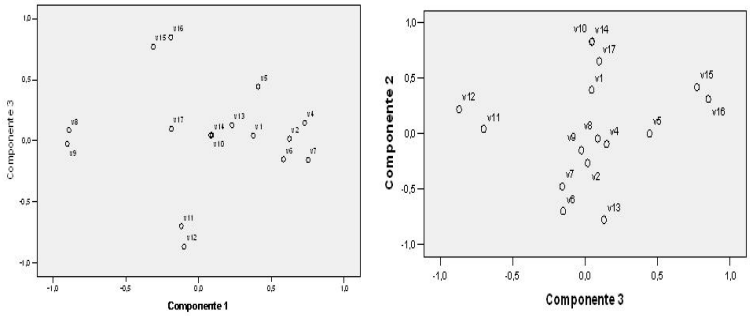


Figura 10.16: Gráficos

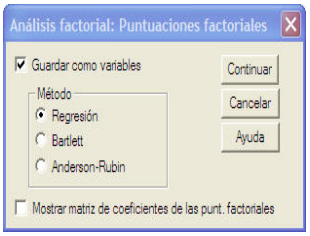


Si estudiamos por ejemplo el gráfico de las componentes 2 y 3, vemos que las variables Gasto per Capita y PIB en salud están correladas entre si y positivamente con el factor tres, frente a Población protegida por Hospital y Ambulatorio, por lo que este tercer eje marcaría la separación entre el gasto y la población protegida.

10.2.4. Puntuaciones Factoriales

Finalmente, si escogemos el *Botón Puntuaciones* y seleccionamos *Guardar como variables*:

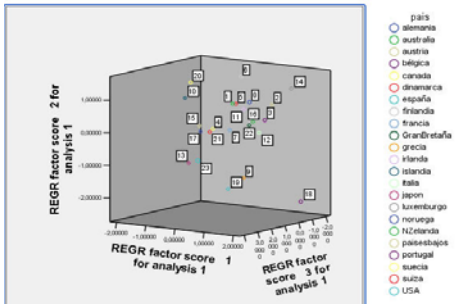
Figura 10.17: Puntuaciones factoriales



Obtenemos las puntuaciones factoriales, que consisten en reducir el conjunto de variables de cada individuo al conjunto de factores o dicho en modo matemático, las puntuaciones

de cada individuo sobre el conjunto de factores extraído. Estas puntuaciones se añaden en el Fichero de Datos, con los nombres FAC1_1, FAC1_2, FAC1_3. De tal manera que solo tendremos que representar estas tres variables (Menú Gráficos; Dispersión; Dispersión 3D; Introducir las tres variables nuevas en los ejes y establecer marcas por País). Resultando el siguiente gráfico: pudiendo sacar las conclusiones referentes a los países, como que Islandia tiene un

Figura 10.18: Gráfico puntuaciones factoriales



comportamiento parecido a Suecia, o que España es similar a Grecia y USA con Japón.

10.3. Ejercicios obligatorios

10.3.1. Ejercicio primero

Una empresa dedicada al diseño de automóviles desea estudiar cuales son los deseos del comprador. Para ello realiza una encuesta a 20 individuos preguntándoles 10 características de sus productos que valoran de 1 a 5. Los datos están en el fichero **factorial2**. Intentar encontrar los factores de comportamiento latentes en los encuestados, identificar los factores numérica y gráficamente. ¿Sería factible alguna rotación? Obtener las puntuaciones factoriales y su gráfico tridimensional.

10.3.2. Ejercicio segundo

Se estudian 100 individuos para comprobar la idea que los consumidores tienen sobre una empresa. Para ello se estudian siete variables sobre la empresa. Se desea realizar un análisis factorial para intentar reducir la dimensión de 7 a menos variables. Realizar el análisis factorial del fichero **Factorial3**.