

**ESTADÍSTICA**  
**Prácticas de ordenador con SPSS para Windows**  
**Licenciatura en Documentación**  
**Curso Académico 2001/2002**  
**Departamento de Estadística e Investigación Operativa**

**Profesora: Dra. Josefa Marín Fernández**

**Índice de contenidos**

<b>1. Normas generales de SPSS. Gestión de ficheros, casos y variables</b>	<b>3</b>
1.1. Introducción . . . . .	3
1.2. Barra de menús . . . . .	4
1.3. Tipos de ventanas . . . . .	4
1.4. Presentación de nuestro ejemplo . . . . .	6
1.5. Definición de las variables . . . . .	6
1.6. Definición de las variables de nuestro ejemplo . . . . .	10
1.7. Grabación de ficheros de datos . . . . .	13
1.8. Introducción de los datos a través del teclado . . . . .	14
1.8.1. Introducción de datos por casillas . . . . .	14
1.8.2. Introducción de datos por filas (casos o sujetos) . . . . .	15
1.8.3. Introducción de datos por columnas (variables) . . . . .	15
1.8.4. Introducción de datos en un área delimitada de casos y variables . . . . .	15
1.9. Introducción de los datos del ejemplo . . . . .	16
1.10. Lectura de ficheros de datos propios de SPSS . . . . .	16
1.11. Edición y modificación de los datos . . . . .	17
1.12. Impresión de los datos . . . . .	21
1.13. Importación de ficheros de datos . . . . .	21
1.14. Transformación de variables . . . . .	21
1.14.1. Modificación sin utilizar el condicional . . . . .	22
1.14.2. Modificación condicional de variables . . . . .	25
1.14.3. Recodificación de variables . . . . .	26
1.14.4. Recodificación automática de variables . . . . .	28
1.14.5. Ordenación por rangos . . . . .	29
1.14.6. Otras modificaciones . . . . .	30
1.15. Ponderación de los casos . . . . .	30
1.16. Ordenación de los casos . . . . .	32
<b>2. Estadística descriptiva unidimensional</b>	<b>34</b>
2.1. Distribución de frecuencias . . . . .	34
2.2. Observaciones sobre la presentación, modificación y grabación de los resultados en la ventana del visor de SPSS . . . . .	35
2.2.1. Cómo controlar la visualización de la información sobre variables y valores de datos en las tablas pivote y en los titulares . . . . .	35
2.2.2. Cómo establecer el aspecto de las tablas pivote por defecto . . . . .	35
2.2.3. Cómo cambiar las especificaciones del visor de resultados de SPSS por defecto . . . . .	36
2.2.4. Cómo guardar y recuperar una ventana del visor de resultados de SPSS . . . . .	37
2.3. Medidas descriptivas de los datos . . . . .	37
2.4. Representaciones gráficas unidimensionales . . . . .	41
2.4.1. Cómo cambiar las especificaciones de las representaciones gráficas por defecto . . . . .	41
2.4.2. Creación de diagramas de barras, líneas, áreas y sectores . . . . .	43

2.4.3.	Edición y modificación de diagramas de barras, líneas, áreas y sectores . . . . .	49
2.4.4.	Creación de histogramas . . . . .	52
2.4.5.	Edición y modificación de histogramas . . . . .	54
2.4.6.	Otras funciones sintéticas . . . . .	55
2.4.7.	Cuadro de diálogo de Títulos . . . . .	56
2.4.8.	Cuadro de diálogo de Opciones . . . . .	57
2.4.9.	Sobre la Plantilla . . . . .	57
<b>3.</b>	<b>Probabilidad. Variables aleatorias</b>	<b>58</b>
3.1.	Función de distribución (probabilidad acumulada) . . . . .	58
3.2.	Inversa de la función de distribución (percentiles) . . . . .	60
<b>4.</b>	<b>Contrastes paramétricos</b>	<b>63</b>
4.1.	Contraste para una media. Intervalo de confianza . . . . .	63
4.2.	Comparación de dos medias con muestras independientes . . . . .	64
4.2.1.	Ejemplo de edición de unos datos para hacer la comparación de dos medias con muestras independientes	66
4.3.	Comparación de dos medias con muestras relacionadas . . . . .	67
<b>5.</b>	<b>Contrastes no paramétricos</b>	<b>70</b>
5.1.	Contraste sobre una proporción . . . . .	70
5.2.	Contraste $\chi^2$ de Pearson sobre proporciones . . . . .	70
5.3.	Contraste $\chi^2$ sobre independencia de dos variables . . . . .	72
5.3.1.	Ejemplo de edición de una tabla de contingencia en el editor de datos . . . . .	76
5.4.	Contraste de las rachas sobre aleatoriedad de la muestra . . . . .	77
5.5.	Contraste de Kolmogorov–Smirnov sobre bondad de ajuste. Normalidad . . . . .	78
5.6.	Contraste de homogeneidad con 2 muestras independientes . . . . .	78
5.7.	Contraste de homogeneidad con más de dos muestras independientes . . . . .	80
5.8.	Contraste de homogeneidad con 2 muestras relacionadas . . . . .	80
5.9.	Contraste de homogeneidad con más de dos muestras relacionadas . . . . .	82

# 1. Normas generales de SPSS. Gestión de ficheros, casos y variables

## 1.1. Introducción

Las prácticas se van a realizar con el programa *SPSS para Windows*, versión 9.0 (en español). Para ejecutar el programa debemos seleccionar Inicio ⇒ Programas ⇒ SPSS 9.0 para Windows. La primera vez que se ejecuta el programa aparece la ventana de la Figura 1, en la que es conveniente marcar la opción **No volver a mostrar este cuadro de diálogo**. Así, cuando volvamos a

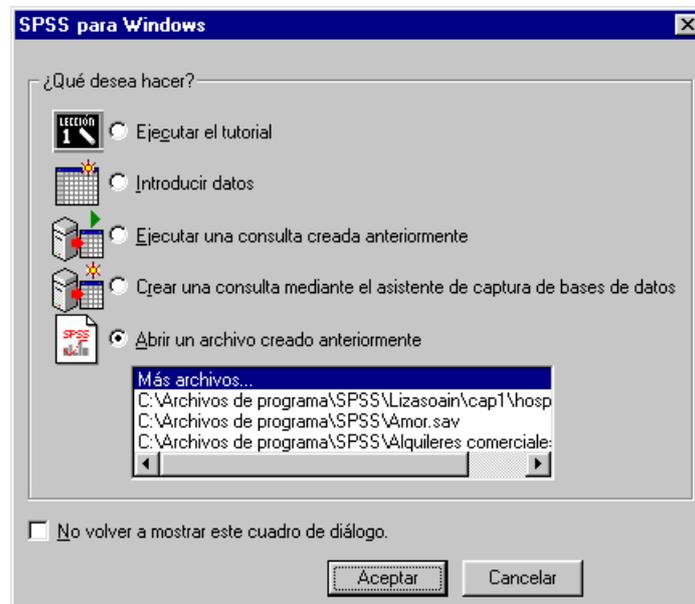


Figura 1: Ventana inicial del programa SPSS 9.0 para Windows

ejecutar el programa *SPSS para Windows* lo primero que aparecerá será la ventana básica de esta aplicación, o para ser más precisos, la *ventana del editor de datos* (Figura 2). Como en cualquier otra aplicación Windows, la *ventana del editor de*

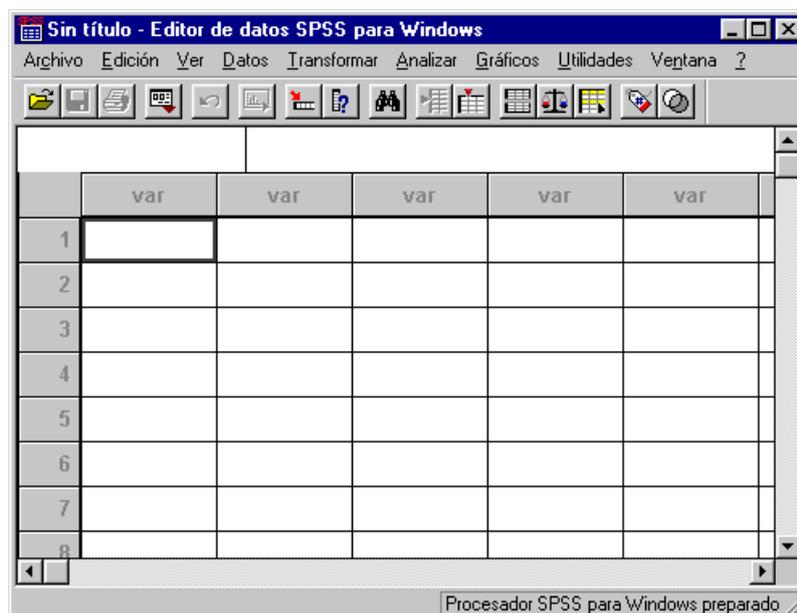


Figura 2: Ventana del editor de datos de SPSS

*datos* puede modificarse en cuanto al tamaño y a la disposición de sus elementos. Posteriormente veremos el resto de las ventanas de la aplicación *SPSS para Windows*; pero, por ahora, vamos a centrarnos en ésta. Se trata de una ventana típica de una aplicación Windows que, de arriba a abajo, consta de los siguientes elementos:

- En la primera línea aparece la **barra de título** con el menú de control, el nombre de la ventana y los botones de minimizar, maximizar y cerrar.
- En la segunda línea está la **barra de menú** con los 10 menús que luego comentaremos.

- La tercera línea es la **barra de herramientas** donde, mediante botones con iconos, se representan algunas de las operaciones más habituales. Si pasamos el puntero del ratón por cualquiera de ellos, aparecerá en la pantalla un texto indicando la función que se activa. Esta barra es personalizable.
- Después aparece la **línea de edición de datos** que, a su vez, está dividida en dos partes, como posteriormente veremos.
- El grueso de la ventana está ocupado por la **matriz de datos** que aparece flanqueada por una primera fila en gris con los encabezamientos de las variables y por una primera columna, también en gris, con los números de los casos.
- La última línea de la ventana (de arriba a abajo) es la **barra de estado**, en la que el sistema proporciona diversos mensajes. Esta barra está dividida en varias zonas de avisos. La primera está reservada para informaciones de tipo general y en la segunda aparece lo relativo al procesador de **SPSS**. Cuando no hay operación en marcha aparece el mensaje *Procesador SPSS para Windows preparado*. En caso contrario aparecerá el nombre del comando que en ese momento se esté ejecutando e información adicional como, por ejemplo, el número de casos procesados.

## 1.2. Barra de menús

A continuación se da un resumen de lo que se puede encontrar en la **barra de menús**:

**Archivo:** Mediante este menú se pueden abrir, crear o grabar los diferentes ficheros que **SPSS** emplea, ya sean de datos, instrucciones, resultados o procesos. Igualmente, es posible controlar las tareas de impresión.

**Edición:** Permite realizar las tareas habituales de edición: modificar, borrar, copiar, pegar, seleccionar, etc.

**Ver:** Permite controlar diversos parámetros de visualización en pantalla.

**Datos:** Este menú permite definir variables, así como efectuar modificaciones en los ficheros de datos: seleccionar, añadir, ponderar, etc.

**Transformar:** Aquí se encuentran todas las opciones relativas a la modificación y generación de nuevas variables. Las funciones de este menú y del anterior son temporales y sólo están en vigor durante la sesión, sin que afecten al fichero original de datos. Si se quieren convertir en permanentes, deben grabarse los cambios.

**Analizar:** Mediante este menú se accede a los diferentes análisis estadísticos que se pueden realizar con los datos.

**Gráficos:** Permite la creación y edición de diversos tipos de gráficos de alta resolución. Algunos de ellos son también accesibles a través de determinadas técnicas estadísticas.

**Utilidades:** Entre otras, posibilita mostrar información sobre los ficheros de **SPSS**, las variables o el tratamiento de conjuntos de variables.

**Ventana:** Dispone de las funciones habituales para controlar las ventanas.

**?:** Proporciona ayuda al usuario en el formato típico de Windows.

Para salir del programa debe seleccionarse la opción **Archivo** ⇒ **Salir**. Los usuarios de *Windows 95*, *Windows 98* ó *Windows NT* pueden abandonar el programa pulsando el botón de la esquina superior derecha de la ventana (botón ).

## 1.3. Tipos de ventanas

Existen diversos tipos de ventanas en **SPSS**:

**Editor de datos.** Esta ventana muestra el contenido del archivo de datos. Con el *Editor de datos* se pueden crear nuevos archivos de datos o modificar los existentes. La ventana del *Editor de datos* se abre automáticamente cuando se inicia una sesión de **SPSS**. No se puede tener más de un archivo de datos abierto al mismo tiempo.

**Visor.** Todas las tablas, los gráficos y los resultados estadísticos se muestran en el *Visor* (ver la Figura 3). Se pueden editar los resultados y guardarlos para utilizarlos posteriormente. La ventana del *Visor* se abre automáticamente la primera vez que se ejecuta un procedimiento que genera resultados.

Se puede utilizar el *Visor* para:

- Examinar los resultados.
- Mostrar u ocultar tablas y gráficos seleccionados.
- Cambiar el orden de presentación de los resultados moviendo los elementos seleccionados.
- Mover elementos entre el visor y otras aplicaciones.

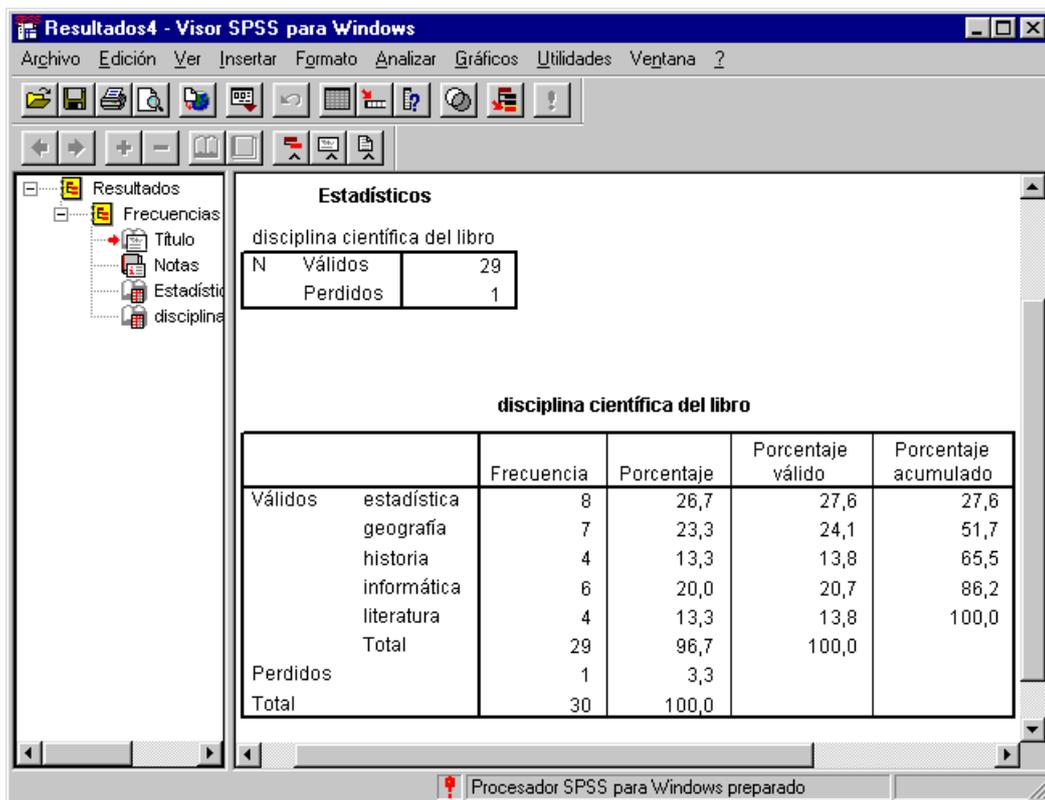


Figura 3: Ejemplo de unos resultados en el Visor de SPSS

El *Visor* se divide en dos paneles:

1. El panel izquierdo muestra una lista de titulares del contenido.
2. El panel derecho contiene tablas estadísticas, gráficos y resultados de texto.

Se pueden utilizar las barras de desplazamiento para examinar los resultados o bien pulsar en un elemento de los titulares para ir directamente a la tabla o gráfico correspondientes.

Se puede pulsar y arrastrar el borde derecho del panel de titulares para cambiar la anchura del mismo.

El panel de titulares proporciona una tabla de contenidos del documento del *Visor*. Se puede utilizar este panel para navegar por los resultados y controlar su presentación. La mayoría de las acciones que hay en dicho panel tienen su efecto correspondiente en el panel de contenidos.

- ◇ Si se selecciona un elemento en el panel de titulares, se seleccionará y mostrará el elemento correspondiente en el panel de contenidos.
- ◇ Si se mueve un elemento en el panel de titulares, también se moverá el elemento correspondiente en el panel de contenidos.
- ◇ Si se contrae la vista de titulares, se ocultarán los resultados de todos los elementos en los niveles contraídos.

**Visor de borrador.** Los resultados pueden mostrarse como texto normal (en lugar de como tablas pivote interactivas) en el *Visor de borrador*.

**Editor de tablas pivote.** Con el *Editor de tablas pivote* es posible modificar los resultados mostrados en este tipo de tablas de diversas maneras. Se puede editar el texto, intercambiar los datos de las filas y las columnas, añadir colores, crear tablas multidimensionales y ocultar o mostrar los resultados de manera selectiva.

**Editor de gráficos.** En ella se pueden modificar los gráficos y diagramas de alta resolución. Es posible cambiar los colores, seleccionar diferentes tipos de fuentes y tamaños, intercambiar los ejes horizontal y vertical, rotar diagramas de dispersión tridimensionales e incluso cambiar el tipo de gráfico.

**Editor de resultados de texto.** Los resultados de texto que no aparecen en las tablas pivote pueden modificarse con el *Editor de resultados de texto*. Se pueden editar los resultados y cambiar las características de las fuentes (tipo, estilo, color y tamaño).

**Editor de sintaxis.** Se pueden pegar las selecciones del cuadro de diálogo en una ventana de sintaxis, donde aparecerán en forma de sintaxis de comandos. A continuación se puede editar esta sintaxis para utilizar las funciones especiales de *SPSS* que no se encuentran disponibles en los cuadros de diálogo. También se pueden guardar los comandos en un archivo para utilizarlos en sesiones de *SPSS* posteriores.

**Editor de procesos.** El procesamiento y la automatización OLE permiten personalizar y automatizar muchas tareas en *SPSS*. Se puede utilizar el *Editor de procesos* para crear y modificar los procesos básicos.

Para acabar lo relativo a las ventanas de *SPSS* veamos una cuestión importante que afecta a las ventanas de resultados y de sintaxis cuando se tienen abiertas más de una de cada tipo.

#### 1.4. Presentación de nuestro ejemplo

Para ilustrar las cuestiones que se van a ir discutiendo a lo largo de las prácticas, utilizaremos el conjunto de datos de la Tabla 1, que se refieren al estudio de diversas variables en una muestra de libros de una biblioteca universitaria. Las variables son: materia (disciplina científica a la que pertenece el libro); precio (en pesetas); entrada (fecha de entrada del libro en la biblioteca); vejez (antigüedad del libro, es decir, número de años transcurridos entre su publicación y la fecha en la que se hace la investigación, con los siguientes resultados: 1=entre 1 y 3 años, 2=entre 4 y 6 años, 3=entre 7 y 9 años, 4=10 años o más); altura (en centímetros); peso (en gramos); préstamo (número de veces que se ha prestado el libro en el último año); y las tres últimas corresponden a la opinión de tres expertos respecto de la encuadernación de los libros (1=muy mala, 2=mala, 3=regular, 4=buena, 5=muy buena).

libro nº	materia	precio	entrada	vejez	altura	peso	préstamo	opinión1	opinión2	opinión3
1	E	5300	03.05.1995	1	20.5	325	32	2	1	1
2	G	3200	06.12.1994	2	25.6	890	18	5	4	5
3	E	12590	25.10.1995	2	20.0	415	9	4	5	5
4	I	7500	03.09.1994	2	20.0	400	15	2	3	4
5	L	890	12.03.1996	1	18.8	515	2	3	3	3
6	I	11550	26.06.1996	2	25.0	650	36	1	5	3
7	H	3250	15.10.1995	1	30.6	790	5	3	3	1
8	H	4000	26.11.1994	4	28.9	890	9	1	3	2
9	L	19500	07.07.1994	1	20.1	320	7	4	2	1
10	I	5500	04.05.1995	1	20.0	420	12	2	5	3
11	E	8700	06.10.1995	3	23.5	620	42	3	2	2
12	E	16300	12.12.1996	2	23.5	720	20	4	3	4
13	G	9650	12.05.1995	3	30.9	850	10	3	3	5
14	I	2500	13.11.1994	2	20.5	620	9	1	3	2
15	E	8750	15.02.1996	1	22.0	475	15	2	5	5
16	L	1250	01.02.1994	1	20.0	320	1	4	4	3
17	L	1250	11.11.1995	1	21.0	380	15	4	3	5
18	G	1250	07.06.1996	2	35.9	790	16	5	4	4
19	G	13750	23.04.1996	1	33.0	780	9	5	3	5
20	I	8100	21.10.1996	2	20.0	500	16	1	2	4
21	E	7500	03.03.1995	2	25.9	550	11	2	5	4
22	G	15850	11.08.1996	1	25.0	820	8	2	3	3
23	H	23000	04.04.1995	1	19.8	360	4	3	2	4
24	H	3950	23.09.1996	4	25.3	490	4	4	1	4
25		2100	15.12.1994	4	22.8	480	2	4	2	4
26	E	6900	06.09.1995	1	19.8	390	23	1	3	5
27	E	18500	12.03.1996	1	20.0	520	16	2	4	3
28	G	10200	08.05.1996	1	20.0	650	13	1	2	3
29	I	4800	04.10.1995	2	20.0	690	20	5	5	4
30	G	3500	06.09.1996	1	36.2	785	9	5	2	5

Cuadro 1: Datos que vamos a introducir y grabar con el nombre LIBROS.SAV

#### 1.5. Definición de las variables

Antes de introducir los datos, tenemos que definir las variables. Para explicar este apartado vamos a hacer la definición de la primera de las variables de nuestro ejemplo. Esta primera variable es la disciplina científica a la que pertenece cada libro y la denominaremos **materia**. Se trata de una variable cualitativa nominal. Para facilitar la introducción de los resultados de

esta variable vamos a codificarla de la siguiente forma: E=estadística, G=geografía, I=informática, L=literatura, H=historia. Para definirla seguimos los siguientes pasos:

- Seleccionamos la primera columna de la matriz de datos haciendo *click* en cualquiera de sus casillas.
- En la barra de menús seleccionamos **Datos** ⇒ **Definir variable**. Entonces aparece el cuadro de diálogo de la Figura 4. Como observamos, la información sobre la definición de variables incluye:



Figura 4: Cuadro de diálogo para definir una variable

- Nombre de variable.
- Tipo de datos (numérico, de cadena, fecha, etc.).
- Las etiquetas descriptivas de variable y de valor.
- Códigos especiales para valores perdidos.
- Nivel de medida.

Se pueden definir variables nuevas o cambiar las definiciones de variables existentes.

- Para los nombres de variable se aplican las siguientes normas:
  - El nombre debe comenzar por una letra. Los demás caracteres pueden ser letras, dígitos, puntos o los símbolos @, #, \_ ó <
  - Los nombres de variable no pueden terminar en punto.
  - Se deben evitar los nombres de variable que terminan con subrayado (para evitar conflictos con las variables creadas automáticamente por algunos procedimientos).
  - La longitud del nombre no debe exceder de los ocho caracteres.
  - No se pueden utilizar espacios en blanco ni caracteres especiales (por ejemplo, !, ?, ' y \*).
  - Cada nombre de variable debe ser único. Los nombres de variable no distinguen mayúsculas de minúsculas.

Nosotros escribiremos **materia** junto a **Nombre de variable**.

- A continuación pulsamos el botón **Tipo**. Entonces aparece el cuadro de diálogo de la Figura 5. Aquí se especifica los tipos de datos de cada variable. Por defecto se asume que todas las variables nuevas son numéricas. Se puede utilizar esta opción para cambiar el tipo de datos. El contenido del cuadro de diálogo *Definir tipo de variable* depende del tipo de datos seleccionado. Para algunos tipos de datos, hay cuadros de texto para el ancho y el número de decimales; para otros, simplemente se puede seleccionar un formato de una lista desplegable de ejemplos. Los tipos de datos disponibles son:

**Numérico:** En este formato, la separación decimal se hace mediante una coma (abajo). En la anchura de los datos cuenta el signo (+ o -), si es que lo vamos a poner, y cuenta la coma de la separación decimal.

**Coma:** Se emplea cuando queremos que la coma sea el separador de los miles y el punto el separador de los decimales.

**Punto:** Se emplea cuando se quiere que el punto sea el separador de los miles y la coma el separador de los decimales.

**Notación científica:** En este formato, cada dato es un número multiplicado por 10 elevado a otro número (positivo o negativo), como por ejemplo 347E-5, que significa 347 multiplicado por 10 elevado a -5; es decir,  $347 \times 10^{-5} = 347 \times \frac{1}{100000} = 0'00347$ .

**Fecha:** Mediante este formato es posible introducir las variables temporales. Comentaremos más sobre este tipo cuando definamos la variable **entrada**.

**Dólar:** A un valor numérico dado se le añade el símbolo del dólar en los diferentes formatos que aparecen en la ventana desplegable.

**Moneda personalizada:** Si a través de la opción **Edición**  $\Rightarrow$  **Opciones**  $\Rightarrow$  **Pestaña de Moneda** de la barra de menús se han creado formatos específicos para este tipo de variables, mediante esta opción se puede elegir uno de ellos.

**Cadena:** Variable cuyos resultados no son numéricos. En su definición debe especificarse únicamente su longitud máxima. Con este tipo de variables, y a diferencia de lo que ocurría con los nombres de las variables, sí hay diferencia entre emplear mayúsculas o minúsculas

En nuestro ejemplo, activamos **Cadena** del cuadro de diálogo (Figura 5), ya que los resultados de esta variable no son numéricos. Al lado de **Caracteres** indicamos un 1, pues los resultados tienen una longitud de un sólo dígito (E, G, I, L ó H). Pulsamos el botón **Continuar** para volver al cuadro inicial de definición de la variable (Figura 4).

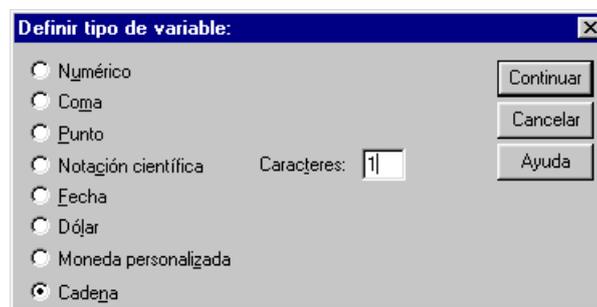


Figura 5: Cuadro de dialogo para definir el tipo de variable

- En el cuadro inicial de definición de la variable (Figura 4) pulsamos el botón **Etiquetas**. La opción *Definir etiquetas* proporciona etiquetas descriptivas de variable y de valor (ver Figura 6). Aunque los nombres de variable sólo pueden tener un máximo de 8 caracteres, las etiquetas de variable pueden tener hasta 256 caracteres. Estas etiquetas descriptivas aparecen en los resultados. A cada valor de una variable se le pueden asignar etiquetas descriptivas de valor. Esto es particularmente útil si el archivo de datos utiliza códigos numéricos para representar categorías no numéricas (por ejemplo, los códigos 1 y 2 para hombre y mujer). Las etiquetas de valor pueden tener una longitud máxima de 60 caracteres. Estas etiquetas no están disponibles para las variables de cadena larga (variables de cadena con más de 8 caracteres). También se pueden modificar o eliminar etiquetas de valor o de variable.

En nuestro ejemplo, podemos escribir *disciplina científica del libro* junto a **Etiqueta de variable**. A continuación, junto a **Valor** escribimos el primero (por ejemplo, E) y junto a **Etiqueta de valor** escribimos lo que le corresponde (estadística). Después pulsamos el botón **Añadir** con lo que la declaración queda incorporada a la lista. Se hace lo mismo con el resto de los valores (ver Figura 6). Una vez finalizado el proceso se pulsa **Continuar**.

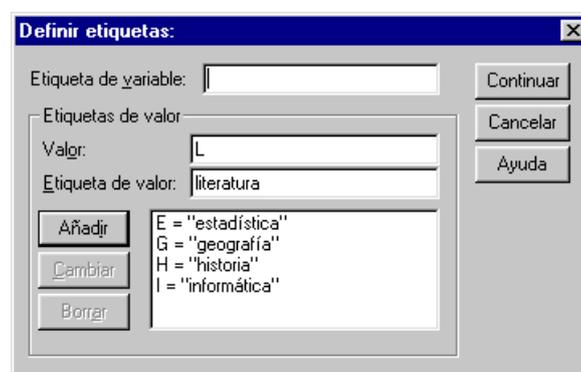


Figura 6: Cuadro de dialogo para definir etiquetas de una variable

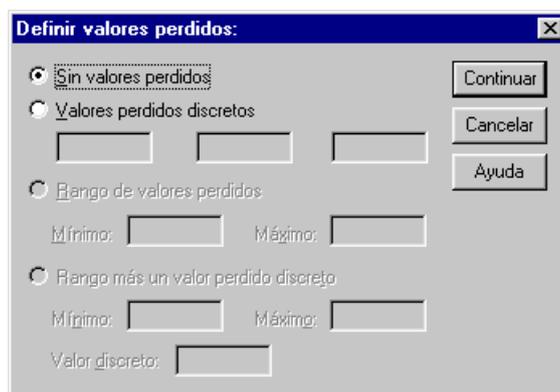


Figura 7: Cuadro de dialogo para definir los valores perdidos de una variable

- En el cuadro inicial de definición de la variable (Figura 4) pulsamos el botón **Valores perdidos** y nos aparece el cuadro de diálogo de la Figura 7. Esta opción permite definir los valores de los datos especificados como perdidos por el usuario. A menudo es útil para saber por qué se pierde información. Por ejemplo, puede desear distinguir los datos perdidos porque un entrevistado se niega a responder, o datos perdidos porque la pregunta no afectaba a dicho entrevistado. Los valores de datos especificados como perdidos por el usuario aparecen marcados para un tratamiento especial y se excluyen de la mayoría de los cálculos.
  - Se pueden introducir hasta tres valores perdidos (individuales) de tipo discreto, un rango de valores perdidos o un rango más un valor de tipo discreto.
  - Sólo pueden especificarse rangos para las variables numéricas.
  - No se pueden definir los valores perdidos para variables de cadena larga (variables de cadena de más de 8 caracteres).
  - Para las variables de cadena, se considera que son válidos todos los valores de cadena, incluidos los valores vacíos o nulos, a no ser que se definan explícitamente como perdidos. Para definir como perdidos los valores vacíos o nulos de una variable de cadena, se escribe un espacio en blanco en uno de los campos de **Valores perdidos discretos**.

En nuestro ejemplo, debemos señalar la opción **Valores perdidos discretos** y dentro del primer campo escribir un espacio en blanco (puesto que hay un valor perdido de dicha variable, que se ha señalado en la Tabla 1 mediante un espacio en blanco). A continuación debemos pulsar el botón **Continuar**.

- Si en el cuadro inicial de definición de la variable (Figura 4) pulsásemos **Aceptar**, la variable **materia** quedaría definida y activada en la matriz, pero dada la longitud declarada (1), el nombre aparecería truncado. Para solucionar esto pulsamos el botón **Formato de columna**. Entonces aparece un cuadro de diálogo como el de la Figura 8. La opción *Definir formato*

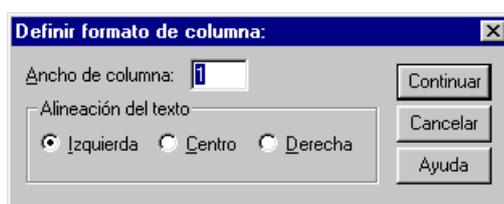


Figura 8: Cuadro de dialogo para definir el formato de la columna correspondiente a una variable

*de columna* controla el ancho de columnas en el *Editor de datos* y la alineación de los valores de datos. Los anchos de columna también se pueden cambiar en el *Editor de datos* pulsando y arrastrando los bordes de la columna. Los formatos de columna afectan sólo a la presentación de valores en el *Editor de datos*. Al cambiar el ancho de columna no se cambia el ancho definido de una variable. Si el ancho real de un valor es más ancho que la columna, aparecerán asteriscos (\*) en la ventana del *Editor de datos*.

En nuestro ejemplo vemos cómo la anchura está definida en 1 carácter correspondiente a la longitud de los resultados de la variable. Ahora se modifica su valor incrementándolo al número de caracteres del nombre de la variable, que es 7. También podemos seleccionar **Derecha** en **Alineación del texto**. Después pulsamos el botón **Continuar** y el control vuelve al cuadro inicial de definición de la variable (Figura 4).

- En este cuadro de diálogo se ha de pulsar el botón **Aceptar** ya que el proceso de definición de la variable **materia** ha terminado. Como resultado, la variable ha dejado de ser potencial y aparece activada en el *Editor de datos* (Figura 9).

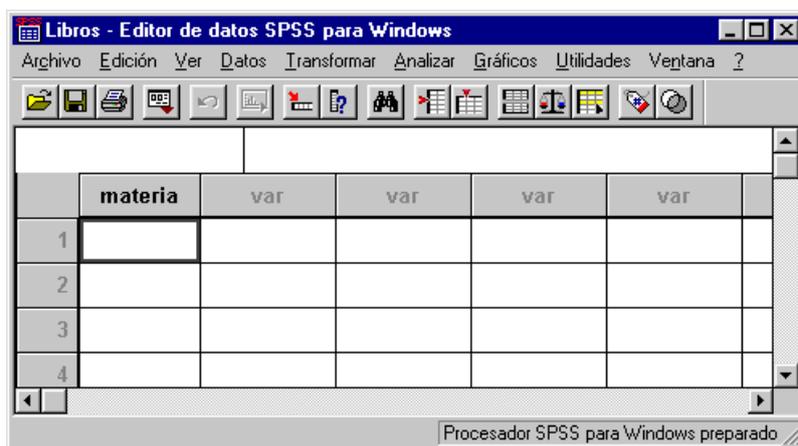


Figura 9: Ventana del editor de datos con la variable materia definida y activada

## 1.6. Definición de las variables de nuestro ejemplo

El procedimiento descrito anteriormente debe realizarse con el resto de las variables de nuestro ejemplo. Como es similar en todas ellas, a continuación sólo vamos a presentar lo que de específico tienen algunas.

En una primera fase vamos a dejar la variable **precio** sin definir, para aprender posteriormente a añadir una variable.

Veamos lo relativo a la variable *fecha de entrada del libro en la biblioteca*, que vamos a denominar **entrada**:

- ◇ En primer lugar, para abrir el cuadro de diálogo *Definir variable* (Figura 4), además de lo visto, podemos optar por otros procedimientos: colocar el puntero del ratón en el encabezamiento de la segunda columna y hacer doble *clic*; o desde el mismo lugar pulsar el botón auxiliar del ratón (el derecho habitualmente) y en el menú contextual que aparece activar la opción **Definir variable**.
- ◇ Junto a **Nombre de variable** (Figura 4) escribimos **entrada**.
- ◇ En el cuadro de diálogo relativo al tipo de la variable (Figura 5), una vez seleccionada la opción correspondiente al tipo **Fecha**, de la ventana que se abre (Figura 10) se selecciona el formato más adecuado. Nosotros seleccionaremos **dd.mm.aaaa** que significa día.mes.año con cuatro dígitos. Luego pulsamos **Continuar**.



Figura 10: Cuadro de diálogo para definir el tipo fecha

- ◇ En cuanto a las variables de este tipo, hay que tener en cuenta varias cuestiones:
  - Las variables de fecha creadas con *Definir fechas* no deben confundirse con las variables con formato de fecha, que se definen en *Definir variable* (como en nuestro ejemplo). Las variables de fecha se emplean para establecer la periodicidad de los datos de serie temporal; mientras que las variables con formato de fecha representan fechas y horas mostradas en varios formatos de fecha y hora. Las variables de fecha son números enteros sencillos que representan el número de días, semanas, horas, etc., a partir de un punto inicial especificado por el usuario. Internamente, la mayoría de las variables con formato de fecha se almacenan como el número de segundos transcurridos desde el 15 de octubre de 1582.

**Observación** ¿Por qué la referencia es la fecha 15 de octubre de 1582?

El actual calendario, aunque retocado en 1582, fue ideado en el año 532 d.C. por el monje y astrónomo Dionisio "el diminuto". Dionisio calculó (aunque se equivocó en, al menos, 4 años) que Jesucristo nació el 25 de diciembre del año 753 A.U.C. ("ab urbe condita" o desde la fundación del Imperio Romano). Dionisio hizo que el 1 de enero del año 1 (no del año cero) coincidiese con el 1 de enero del año 754 A.U.C. (no el nacimiento de Jesús sino 7 días después, que coincidía con la fiesta de la Circuncisión).

Cuando Dionisio ideó nuestro actual calendario estaba vigente el calendario Juliano. Este calendario se estableció en el año 45 a.C. En él, cada año tenía 365'25 días, por lo que cada 4 años había uno con 366 días (bisiesto). Pero cada año se añadía 11 minutos y "pico" al año solar y por tanto cada 1000 años se añadían 7 días al año solar. Por orden del papa Gregorio XIII se reunió un comité en 1578 para dar solución al desajuste de 10 días que ya existía entre el año civil y el año solar. El papa promulgó las nuevas reglas en una bula papal el día 24 de febrero de 1582. Desaparecieron del calendario del 5 al 14 de octubre de 1582 (al día 4 de octubre de 1582 le siguió el día 15 de octubre de 1582). A partir de ese 15 de octubre de 1582 no ha habido más reformas en el calendario. Este calendario, llamado Gregoriano, ya estableció que los años tienen 365'2422 días. También estableció que cada 4 años hay uno bisiesto menos para los que sean divisibles por 100, a excepción de los años que sean divisibles por 400. El año Gregoriano se aparta del solar sólo por 25'96 segundos.

- Los valores de año de dos dígitos de las variables con formato de fecha del editor de datos poseen un valor de siglo basado en los fijados en **Edición** ⇒ **Opciones** ⇒ **Pestaña de Datos** de la barra de menús (ver Figura 11). El valor fijado por defecto es el de 1900 a 1999. Los valores automáticos se basan en el año actual, con principio 69 años antes y final 30 años después de esa fecha. Por ejemplo, si en el año 1999 se introduce una fecha como 10-10-29, se almacena internamente como el 10-10-2029, y la fecha 10-10-65 se almacena como el 10-10-1965. Si se tiene alguna duda, es conveniente indicar siempre el año con cuatro dígitos.



Figura 11: Pestaña de Datos, dentro de la opción Edición ⇒ Opciones

- ◇ Como el nombre que se le ha dado a la variable podría no ser suficientemente explícito, deberíamos pulsar el botón **Etiquetas** para declarar la **Etiqueta de variable** como *fecha de entrada del libro*. Después pulsamos **Continuar**.
- ◇ Aquí no es preciso pulsar **Formato de columna**, dado que al seleccionar el formato de fecha **dd.mm.aaaa**, el ancho de la columna implícito será de 10 (los puntos también cuentan).
- ◇ Así la variable **entrada** ha sido completamente definida, por lo que se pulsa **Aceptar** para finalizar el proceso.

Pasamos ahora a la definición de la variable *antigüedad del libro* (número de años transcurridos entre su publicación y la fecha en la que se hace la investigación). A esta variable la denominaremos **vejez**, debido a que la palabra *antigüedad* tiene más de 8 caracteres.

- ♣ En el cuadro *Definir variable* (Figura 4) escribimos **vejez** junto a **Nombre de variable** y señalamos **Ordinal** en **Medida** (ya que la variable es cualitativa ordinal).
- ♣ En el cuadro *Definir tipo de variable* (Figura 5), que se abre cuando pulsamos **Tipo**, debemos seleccionar **Númérico**. En **Ancho** escribimos 1, y en **Decimales** escribimos 0.
- ♣ En el cuadro *Definir etiquetas* (Figura 6), que se abre cuando pulsamos **Etiquetas**, escribiremos *años transcurridos desde la publicación del libro* en **Etiqueta de variable**. Procediendo de la misma forma que se indicó en la variable **materia**, pero obviamente con las etiquetas de la variable **vejez**, en **Etiquetas de valor** de la variable **vejez** haremos las siguientes correspondencias: 1=entre 1 y 3 años, 2=entre 4 y 6 años, 3=entre 7 y 9 años, 4=10 años o más.

- ♣ En el cuadro *Definir formato de columna* (Figura 8), que se abre cuando pulsamos **Formato de columna**, escribiremos 5 en **Ancho de columna**, que es la anchura del nombre de la variable.

A continuación resumiremos lo que hay que hacer para definir la variable **altura**:

- ♡ En el cuadro *Definir variable* escribimos **altura** junto a **Nombre de variable**.
- ♡ En el cuadro *Definir tipo de variable*, si seleccionamos **Coma** entonces la separación de los miles se escribirá con una coma y la separación de los decimales se escribirá con un punto. Si seleccionamos **Punto** entonces la separación de los miles se escribirá con un punto y la separación de los decimales se escribirá con una coma. Como en español la separación de los miles se hace con un punto y la separación decimal se hace con una coma (aunque está arriba), nosotros elegiremos la opción **Punto**. En **Ancho** debemos escribir 4 (pues son dos cifras para la parte entera más una coma más un decimal), y en **Decimales** escribimos 1.
- ♡ En el cuadro *Definir etiquetas* podemos escribir *altura del libro, en centímetros* en **Etiqueta de variable**.
- ♡ En el cuadro *Definir formato de columna* escribiremos 6 en **Ancho de columna** pues ésta es la anchura del nombre de la variable.

Veamos ahora lo correspondiente a la variable **peso**:

- ♠ En el cuadro *Definir variable* escribimos **peso** junto a **Nombre de variable**.
- ♠ En el cuadro *Definir tipo de variable* señalaremos **N Numérico**. Aunque los resultados obtenidos para la variable **peso** sólo tienen tres dígitos, debemos dar la opción de que posteriormente se introduzca un nuevo caso (libro) con un peso mayor que 1000 gramos; por tanto, en **Ancho** debemos escribir 4. En **Decimales** escribimos 0.
- ♠ En el cuadro *Definir etiquetas* podemos escribir *peso del libro, en gramos* en **Etiqueta de variable**.
- ♠ En el cuadro *Definir formato de columna* escribiremos 4 en **Ancho de columna**.

El resumen de lo correspondiente a la definición de la variable *número de veces que se ha prestado el libro en el último año* es:

- En el cuadro *Definir variable* escribimos **préstamo** junto a **Nombre de variable**.
- En el cuadro *Definir tipo de variable* señalaremos **N Numérico**. Aunque los resultados obtenidos para la variable sólo tienen dos dígitos, debemos dar la opción de que posteriormente se introduzca un nuevo caso (libro) que se haya prestado 100 veces o más durante el último año; en consecuencia, en **Ancho** debemos escribir 3. En **Decimales** escribimos 0.
- En el cuadro *Definir etiquetas* podemos escribir *número de veces que se ha prestado el libro en el último año* en **Etiqueta de variable**.
- En el cuadro *Definir formato de columna* escribiremos 8 en **Ancho de columna**.

Y ahora sólo nos quedan las tres variables relativas a la *opinión* de tres expertos sobre la encuadernación de los libros. Este es el típico caso en que un grupo de variables comparten una misma estructura, por lo que es posible definir sus características de forma conjunta (excepto el nombre, por supuesto) empleando las *plantillas*.

- En primer lugar se definen, una a una, las tres variables, cuyos nombres respectivos serán: **opinión1**, **opinión2** y **opinión3**. En cada una de ellas hay que especificar el **ordinal** como nivel de **Medida**. Acabado el proceso de definición sencilla de estas tres variables, las tres columnas están ya activas en la matriz de datos.
- Se seleccionan las tres variables, situando el puntero en el encabezamiento de la primera y, sin soltar el botón izquierdo del ratón, se arrastra hasta llegar a la última.
- Del menú **Datos** se selecciona **Plantillas**, lo que origina que se abra el cuadro de diálogo de la Figura 12. De este cuadro de diálogo seleccionamos en **Plantilla** la opción **PREDETERMINADA** (o **DEFAULT**) y pulsamos el botón **Definir**.
- En el cuadro de diálogo que aparece (Figura 13) comenzamos la declaración de los diferentes apartados. Los cuadros son los mismos que para las variables, así que hay poco de nuevo en esta fase.
- Con respecto al **Tipo**, especificamos variable numérica con extensión de 1 carácter y cero dígitos decimales.
- El proceso de etiquetado de los valores es idéntico al realizado con la variable **materia**, pero lógicamente hay que poner las etiquetas de estas tres nuevas variables, que corresponden a la opinión de tres expertos respecto de la encuadernación de los libros. Estas etiquetas de los valores son: 1=muy mala, 2=mala, 3=regular, 4=buena, 5=muy buena.

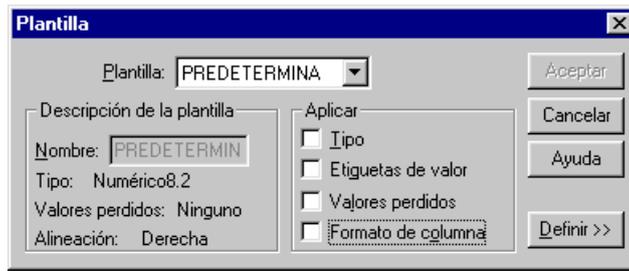


Figura 12: Cuadro de diálogo para hacer plantillas de variables



Figura 13: Cuadro de diálogo para definir una plantilla

- A continuación se le da nombre a la plantilla; por ejemplo, *OPINIÓN*, tecleándolo en el recuadro que hay junto a **Nombre** en el apartado **Descripción de la plantilla** del cuadro de diálogo de la Figura 13.
- En el apartado **Aplicar** del cuadro de diálogo de la Figura 13 hay que activar las características de la plantilla que queremos que se apliquen a las variables seleccionadas. En nuestro caso seleccionamos **Tipo** y **Etiquetas de valor**.
- Se pulsa el botón **Añadir** para que la nueva plantilla que acabamos de definir se incorpore a la lista.
- Por último, pulsamos **Aceptar** para finalizar el proceso y aplicar las especificaciones en las tres variables seleccionadas.

En todo momento se pueden cambiar las especificaciones declaradas en la definición de cualquier variable, y para ello basta con reiniciar el proceso de definición de las variables cuyas características se quieran alterar. Siempre es posible, por ejemplo, aumentar la longitud de una variable, introducir nuevos códigos para los valores ausentes o perdidos, añadir o modificar etiquetas, etc.

Pero si los datos ya han sido introducidos, y sobre todo si son muchos, hay otras serie de posibles modificaciones que son menos recomendables, como cambiar el tipo asignado a las variables, disminuir su longitud o modificar valores ausentes.

Con todo esto, ya disponemos de la matriz completamente definida. Se dice entonces que el *diccionario* ya está incorporado al sistema, y sólo queda introducir los datos.

## 1.7. Grabación de ficheros de datos

Aunque todavía no hemos introducido los datos correspondientes a las variables que hemos definido, si no queremos perder toda la información referente a las variables (debido a un corte de electricidad o cualquier otra causa), podemos grabar ya el que será nuestro fichero de datos denominado **Libros.sav**

Si de la barra de menús se selecciona la opción **Archivo** se observa que hay dos opciones para la grabación de ficheros de datos: **Guardar** y **Guardar como**.

Al igual que en otras muchas aplicaciones Windows, la opción **Guardar** sirve para grabar un fichero de datos que ya estaba previamente grabado, con las modificaciones introducidas desde la última vez que se grabó. Las modificaciones que se introduzcan en una sesión en un fichero de datos sólo tienen efecto durante esa sesión, y al finalizarla o cerrar el programa se pierden, a no ser que se graben mediante esta opción. En cualquier caso, al cerrar el programa *SPSS*, si se han introducido modificaciones en el fichero de datos, el sistema siempre envía una pregunta al usuario con objeto de confirmar si se quieren grabar o no los cambios realizados.

La opción **Guardar como** se utiliza al ir a grabar por primera vez un fichero o al ir a modificar algunos de sus parámetros de grabación. Al activarla aparece un cuadro de diálogo como el de la Figura 14. En el mismo debemos especificar, mediante los procedimientos habituales, el nombre del fichero, la unidad y la carpeta de la misma en la que queremos que se grabe. Con

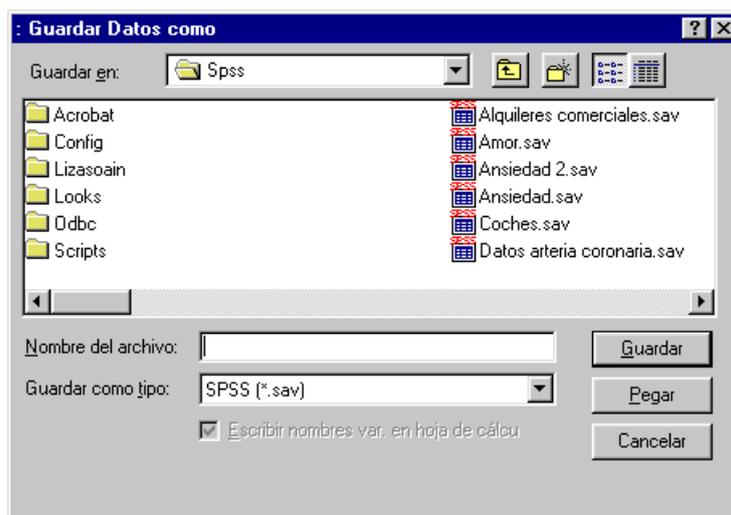


Figura 14: Cuadro de diálogo para grabar un fichero de datos

respecto a la posibilidad de grabar el fichero con un tipo distinto al de **SPSS para Windows**, esto lo veremos posteriormente. En consecuencia, en *Guardar como tipo* dejamos inalterada la opción seleccionada por defecto que es la de un fichero de datos de **SPSS para Windows** (.sav). En el cuadro de diálogo de la Figura 14 vemos que también aparece una opción desactivada, que es la que permite optar entre grabar o no los nombres de las variables cuando el formato de grabación elegido es el correspondiente a las hojas de cálculo.

Ha llegado pues el momento de que procedamos a grabar nuestro fichero de datos con el nombre **Libros.sav**

## 1.8. Introducción de los datos a través del teclado

El resultado de todo el proceso anterior es la matriz que aparece en la Figura 15. En ella se observa que las variables definidas están activas, y seleccionada la casilla correspondiente al primer sujeto y primera variable.



Figura 15: Ventana del editor de datos con las variables definidas

En este punto podemos comenzar con la introducción de los datos. Tan entretenida tarea se puede realizar prácticamente de cualquier forma y orden: sujeto a sujeto, variable a variable, seleccionando una porción de matriz, etc. Pero, como es lógico, no todos los procedimientos son iguales. Así que lo mejor es ver cada uno de ellos.

### 1.8.1. Introducción de datos por casillas

Para introducir un valor en una casilla cualquiera (independientemente de la fila o columna a la que pertenezca) en primer lugar ésta debe seleccionarse situando el puntero en la misma y haciendo *click*. Si ahora se teclea el valor correspondiente, el mismo aparece en la *línea de edición de datos*, pero todavía no ha sido introducido realmente. Para que ello ocurra podemos hacer una de las tres opciones siguientes:

- Pulsar la tecla **Intro**, con lo que se activa la casilla inmediatamente inferior.

- Pulsar la tecla del tabulador, , con lo que se activa la siguiente casilla de la derecha.
- Situar el puntero y hacer *clik* en cualquier otra casilla.

Si por cualquiera de estos procedimientos de desplazamiento y entrada de datos introducimos un valor fuera de los límites de nuestra matriz, automáticamente se crea un nuevo caso o una nueva variable en función de si el límite traspasado es el de las filas o el de las columnas. Si esto ocurre, el valor escrito se introduce en la casilla en cuestión, y el resto de las pertenecientes a ese nuevo caso o nueva variable recién generada son completadas con valores ausentes definidos por el sistema.

Vamos ahora a practicar introduciendo, por ejemplo, los tres primeros datos de la primera variable (que son E, G y E). Podemos seguir introduciendo más datos con este método sin preocuparnos por si cometemos errores, por que para eso está el apartado dedicado a la edición de los datos. Tampoco nos preocuparemos si creamos más casos o variables de los definidos por que ya lo arreglaremos.

### 1.8.2. Introducción de datos por filas (casos o sujetos)

Primero se activa una fila entera haciendo *clik* en el botón donde aparece el número de la misma. De esta manera el registro entero queda seleccionado y el efecto de las teclas de  y  es el mismo, quedando restringido al movimiento entre columnas para ese caso.

Podemos practicar introduciendo los datos de la primera fila (primer caso; es decir, primer libro).

- El primer dato es el resultado de la variable **materia** para el primer libro observado. Este dato ya estaba escrito y es una E (de “estadística”).
- El segundo dato es la fecha de **entrada** del primer libro que se puede introducir de cualquiera de las siguientes maneras: 3-5-95, 3 5 1995, 3/5/95, 3.5.95, 03-5-95, 03 5 1995, 03/5/95, 03.5.95, 3-05-95, 3 05 1995, 3/05/95, 3.05.95, ..., 03-05-1995, 03 05 1995, 03/05/1995, 03.05.1995; pero independientemente de cuál haya sido el escogido, la apariencia en el editor es el del último (03.05.1995). Como vemos, para las variables de tipo fecha, al introducir los datos se puede utilizar como separador el punto, el guión, el espacio en blanco o la barra inclinada; pero no se puede utilizar, por ejemplo, el punto y coma, ni los dos puntos.
- El tercer dato es la **vejez** del primer libro y hay que escribir un 1 (“entre 1 y 3 años desde su publicación”).
- El cuarto dato es la **altura**, en centímetros, del primer libro, y para introducirla tenemos que teclear necesariamente 20,5 pues no se puede omitir la separación decimal (que hemos obligado a que sea una coma en la parte inferior). Por supuesto, no se debe teclear 205, pues en este caso **SPSS** guardaría como dato el 205,0. Tampoco se puede teclear 20.5 pues hemos elegido el tipo *Punto* dentro del cuadro de diálogo *Definir tipo de variable*. Si hubiésemos elegido el tipo *Numérico*, tampoco se podría teclear 20.5 pues la versión de **SPSS** que estamos utilizando está en español (y, por tanto, la separación decimal se indica, por defecto, con una coma abajo).
- El quinto dato es el **peso**, en gramos, del primer libro, y para introducirlo escribimos simplemente 325.
- El sexto dato de la primera fila es el número de veces que se ha prestado el primer libro durante el último año (**préstamo**) y es 32.
- Los tres últimos datos de la primera fila corresponden a la **opinión** de los tres expertos respecto de la encuadernación de este primer libro. Los datos son 2, 1 y 1.

### 1.8.3. Introducción de datos por columnas (variables)

Análogamente, en este caso, en primer lugar se selecciona la variable haciendo *clik* en el encabezamiento de la misma (el botón con el nombre). Así, la columna entera queda seleccionada y las dos teclas citadas ( y ) sirven para desplazarse una posición hacia abajo.

Podemos practicar ahora introduciendo los 10 primeros casos de la segunda variable.

### 1.8.4. Introducción de datos en un área delimitada de casos y variables

Este procedimiento combina los dos anteriores. Mediante la técnica de *marcar y arrastrar*, con el puntero se selecciona un área de la matriz. Una vez que ésta ha sido activada, con la tecla  nos desplazamos hacia abajo en la primera variable, y al llegar al último caso de la misma se activa la primera casilla de la siguiente variable, y así hasta finalizar con el último caso de la última variable. De esta manera, la entrada de datos se hace por variables pero limitándose a los casos seleccionados.

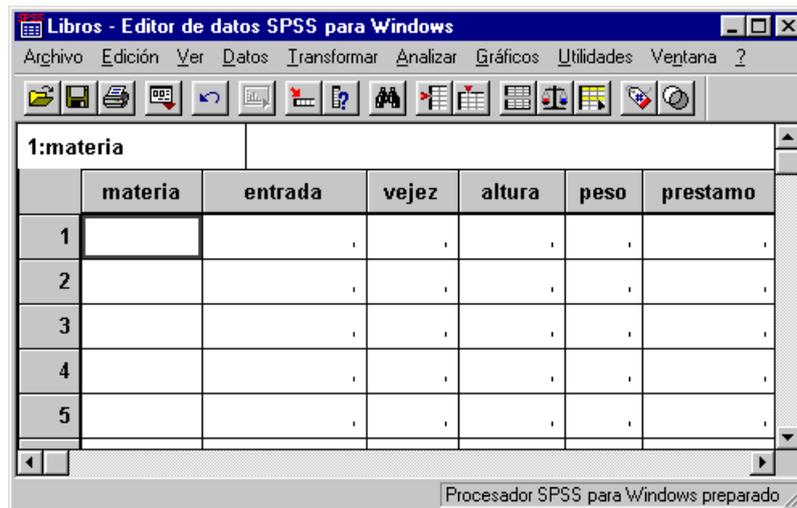
Si la entrada de datos quiere hacerse por casos, pero limitándose a las variables seleccionadas, se usa la tecla  que se va desplazando a la derecha hasta llegar a la última columna del primer caso, para, a continuación, pasar a la primera del segundo, y así hasta finalizar.

## 1.9. Introducción de los datos del ejemplo

Vamos a introducir todos los casos de nuestro ejemplo menos el último (por tanto, 29 casos) con el fin de practicar después la forma de añadir casos.

Antes de introducir los datos de nuestro ejemplo vamos a limpiar la ventana del editor borrando toda la matriz. Para ello, la seleccionamos entera usando la técnica de *señalar y arrastrar* y elegimos la opción **Edición** ⇒ **Eliminar** de la *barra de menús*. Pero antes de hacerlo tengamos en cuenta lo siguiente:

Si comenzamos la selección desde la primera casilla (caso 1, variable 1) hasta la última, al ejecutar la opción **Edición** ⇒ **Eliminar** habremos eliminado todos los valores previamente introducidos; pero tanto las variables como los sujetos (casos) siguen igualmente activos. La prueba de ello es que en todas las casillas de variables numéricas aparecerá la coma indicando que son valores ausentes definidos por el sistema. Como resultado se dispone de una matriz como la de la Figura 16.



	materia	entrada	vejez	altura	peso	prestamo
1	,	,	,	,	,	,
2	,	,	,	,	,	,
3	,	,	,	,	,	,
4	,	,	,	,	,	,
5	,	,	,	,	,	,

Figura 16: Ventana del editor de datos llena de valores ausentes

Por el contrario, si comenzamos la selección en el botón de encabezamiento del primer caso (el marcado con el 1), lo que se selecciona es toda la matriz excepto los encabezamientos de las variables, y el resultado es una matriz sólo definida en lo que se refiere a las variables; el resto está completamente vacío (ver Figura 15).

Señalado esto, procedemos a borrar lo que tengamos en la ventana del editor de datos de la manera que nos guste, y pasamos a introducir los 29 casos. Para ello recomiendo hacer lo siguiente:

1. Recomiendo introducir los datos por columnas (variables), para ir grabando los datos cada vez que introduzcamos cada nueva columna.
2. Recomiendo también, en las variables numéricas (todas menos la primera) utilizar el teclado numérico (el que está a la derecha). Para que este teclado esté activo debe estar encendida la luz que hay junto a *Bloq.Num.*.
3. Recomiendo también introducir las fechas con el separador /, pues éste está en el teclado numérico (por ejemplo, teclearemos 3/5/95).
4. Como en el teclado numérico no está la separación decimal mediante la coma, pero sí está la separación decimal mediante el punto, en este momento (antes de comenzar la introducción de los datos) cambiaremos la definición de la variable altura en cuanto al tipo (ver el apartado 1.5): tipo *Punto* (el punto como separación de los miles y la coma como separador de los decimales) lo cambiaremos por tipo *Coma* (la coma como separación de los miles y el punto como separador decimal).

Una vez hechas estas consideraciones introducimos nuestros datos (sólo hasta la fila número 29). Al final debemos tener la matriz de datos de la Figura 17. No olvidemos grabar ahora el fichero de datos. Después podemos cerrar la sesión de trabajo abandonando el programa con la opción **Archivo** ⇒ **Salir** o pulsando el botón de la esquina superior derecha de la ventana (botón ).

## 1.10. Lectura de ficheros de datos propios de SPSS

Para leer datos grabados mediante *SPSS* hay que elegir **Archivo** ⇒ **Abrir** o pulsar el primer icono de la barra de herramientas. Si se ha trabajado recientemente con el archivo de datos que se quiere abrir, es muy probable que éste se encuentre en la lista de archivos que hay al desplegar el menú **Archivo** (ver Figura 18). Si se encuentra aquí el nombre del fichero, se puede también abrir haciendo *clic* sobre dicho nombre. Para practicar, podemos abrir ahora otro fichero de datos, que no sea **Libros.sav**.

Libros - Editor de datos SPSS para Windows

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1:prestamo

	materia	entrada	vejez	altura	peso	prestamo
1	E	03.05.1995	1	20.5	325	32
2	G	06.12.1994	2	25.6	890	18
3	E	25.10.1995	2	20.0	415	9
4	I	03.09.1994	2	20.0	400	15
5	L	12.03.1996	1	18.8	515	2
6	I	26.06.1996	2	25.0	650	36
7	H	15.10.1995	1	30.6	790	5
8	H	26.11.1994	4	28.9	890	9
9	L	07.07.1994	1	20.1	320	7
10	I	04.05.1995	1	20.0	420	12

Procesador SPSS para Windows preparado

Figura 17: Ventana del editor rellena con los datos del ejemplo

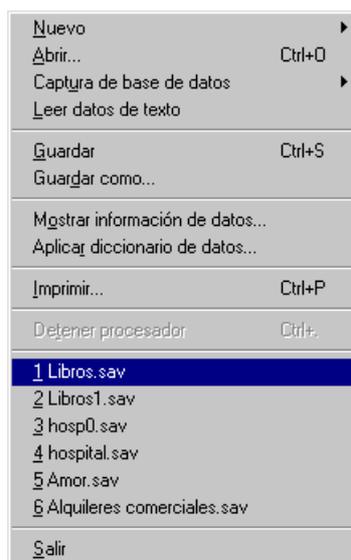


Figura 18: Menú Archivo

### 1.11. Edición y modificación de los datos

En este apartado trataremos sobre cualquier tarea de modificación de los datos de un fichero dado.

- Para **moverse** por la ventana del editor de datos se emplean las teclas de las flechas de movimiento,  $\leftarrow$ ,  $\rightarrow$ ,  $\uparrow$  y  $\downarrow$ . Como ya hemos dicho, la tecla  $\rightleftarrows$  cambia a la casilla contigua de la derecha y la tecla  $\text{Intro}$  cambia a la tecla contigua de abajo. Para ir a la primera variable se pulsa  $\text{Ctrl} + \leftarrow$ . Para ir a la última variable  $\text{Ctrl} + \rightarrow$ . Para ir al principio se pulsa  $\text{Ctrl} + \text{Inicio}$ . Para ir al final,  $\text{Ctrl} + \text{Fin}$ .
- Para **ir a un caso** concreto se elige la opción **Datos**  $\Rightarrow$  **Ir a caso** o se emplea el séptimo icono de la barra de herramientas.
- Para **corregir un dato** se selecciona la casilla correspondiente. El valor en cuestión aparecerá en la *barra de edición de datos*. Se escribe el valor corregido y se pulsa  $\text{Intro}$ .
- Para **seleccionar** un caso (fila) se hace *click* sobre el número del caso. Para seleccionar una variable (columna) se hace *click* sobre el nombre de ella. Para seleccionar un rectángulo de datos se hace *click* en la casilla de la esquina y se arrastra

hasta la otra esquina.

- Para **buscar un dato** de una cierta variable, se selecciona la variable y se elige **Edición** ⇒ **Buscar** o bien se emplea el noveno icono de la barra de herramientas.

Por ejemplo, si buscamos los libros cuya *altura* es igual 25 centímetros veremos que son los numerados con el 6 y el 22.

- Para **insertar un nuevo caso** entre otros ya existentes, nos situamos en el caso siguiente y elegimos **Datos** ⇒ **Insertar caso**, o bien empleamos el décimo icono de la barra de herramientas, o bien pulsamos el botón auxiliar del ratón (el derecho) y elegimos en el menú contextual que aparece (Figura 19) la opción **Insertar caso**.

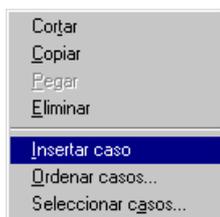


Figura 19: Menú que aparece al seleccionar un caso

Si queremos insertar nuevos casos al final de los ya existentes, basta con introducirlos de la forma indicada en el apartado 1.8. Por ejemplo, vamos a insertar al final el caso (número 30) que nos faltaba.

- Para **ir a una variable** concreta se elige la opción **Utilidades** ⇒ **Variables** y en el cuadro de diálogo que aparece se pulsa el botón **Ir a** (ver Figura 20). También aparece este cuadro de diálogo pulsando el botón del octavo icono de la barra de herramientas.

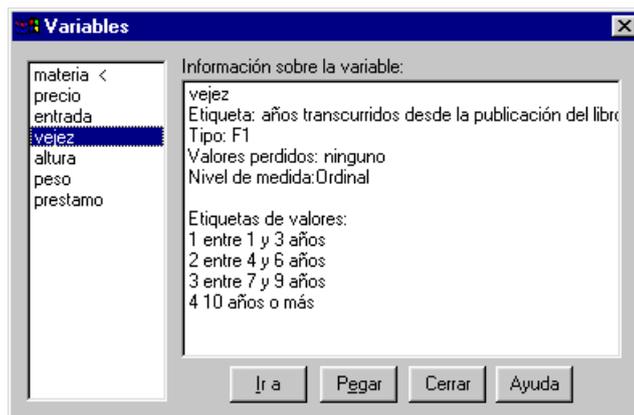


Figura 20: Cuadro que aparece con Utilidades ⇒ Variables

- Para **insertar una variable** entre otras ya existentes, nos situamos en la variable siguiente y elegimos **Datos** ⇒ **Insertar variable**, o bien empleamos el undécimo icono de la barra de herramientas, o bien pulsamos el botón auxiliar del ratón y elegimos en el menú contextual que aparece (Figura 21) la opción **Insertar variable**.

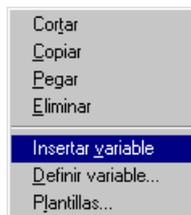


Figura 21: Menú que aparece al seleccionar una variable

Si queremos insertar nuevas variables al final de las ya existentes, basta con definir las e introducir los datos de la forma indicada en los apartados 1.5 y 1.8.3.

Nosotros vamos a insertar la variable **precio** entre las variables **materia** y **entrada**. Para ello:

- Seleccionamos la variable **entrada**.

- Elegimos la opción **Datos** ⇒ **Insertar variable**.
- Una nueva variable es generada, con el tipo y formato preestablecido por el sistema. La columna correspondiente tiene 30 datos ausentes (ver Figura 22).

	materia	var00001	entrada	vejez	altura
1	E	.	03.05.1995	1	20.5
2	G	.	06.12.1994	2	25.6
3	E	.	25.10.1995	2	20.0
4	I	.	03.09.1994	2	20.0
5	L	.	12.03.1996	1	18.8
6	I	.	26.06.1996	2	25.0
7	H	.	15.10.1995	1	30.6
8	H	.	26.11.1994	4	28.9
9	L	.	07.07.1994	1	20.1
10	I	.	04.05.1995	1	20.0

Figura 22: Editor de datos con la variable insertada

- Definimos ahora la variable con los siguientes parámetros (ver apartado 1.5):
  - Nombre de variable: **precio**
  - Etiqueta de variable: *precio del libro, en pesetas*
  - Tipo: *Númérico*, **Ancho: 5**, **Decimales: 0**
  - Formato de columna: **Anchura=6**
- Introducimos ahora los resultados de esta variable de la forma indicada en el apartado 1.8.3.
- Para **cambiar las especificaciones de una variable** se procede de forma análoga a la definición inicial de las variables (ver apartado 1.5). Recordemos que, una vez que hemos introducido los datos, no es conveniente cambiar ciertas especificaciones como el tipo (de cadena a numérico, etc.) o disminuir la longitud. Para comprobar las especificaciones de cada variable se puede revisar el diccionario seleccionando la opción **Utilidades** ⇒ **Variables**. Entonces aparece el cuadro de diálogo de la Figura 20. También aparece este cuadro de diálogo pulsando el botón del octavo icono de la barra de herramientas.

Otras funciones típicas de todos los editores, y por supuesto de éste, son **borrar**, **copiar**, **mover** y **deshacer**. *Mover* consiste en seleccionar, cortar, situarse en otro sitio y pegar. *Copiar* consiste en seleccionar, copiar, situarse en otro sitio y pegar. *Deshacer* es eliminar la última acción de borrar, cortar o pegar. Estas funciones están en el menú **Edición** (ver Figura 23). *Deshacer* también está en el quinto icono de la barra de herramientas.



Figura 23: Menú Edición

Para terminar este apartado diremos que es posible **modificar la apariencia del editor de datos** o su modo de operar. Para ello seleccionamos **Ver** de la barra de menús y hacemos *clic* en la opción deseada del menú que aparece (Figura 24). Estas opciones son:

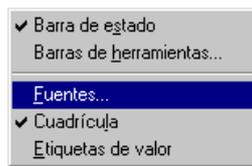


Figura 24: Menú Ver

**Barra de estado:** Sirve para mostrar (cuando está señalado con un ✓) u ocultar (cuando no lo está) la barra de estado del sistema.

**Barras de herramientas:** Sirve para mostrar u ocultar, personalizar y crear nuevas barras de herramientas. En las barras de herramientas puede incluirse cualquier herramienta disponible, incluso la de cualquier acción de menú. Además pueden contener herramientas personalizadas que ejecutan otras aplicaciones, que ejecutan archivos de sintaxis de comandos o archivos de procesos.

**Fuentes:** Si se elige esta opción se abre un cuadro de diálogo similar al de todas las aplicaciones *Windows*, que permite seleccionar las fuentes, estilos y tamaños.

**Cuadrícula:** Sirve para mostrar (cuando está señalado) u ocultar (cuando no lo está) la cuadrícula que separa las diferentes casillas del editor de datos.

**Etiquetas de valor:** Sirve para mostrar (cuando está señalado) u ocultar (cuando no lo está) las etiquetas de las variables en el editor de datos. Por ejemplo, en la Figura 25 se puede ver el aspecto del editor de datos cuando está activada (señalada con ✓) la opción **Etiquetas de valor** y no está activada la opción **Cuadrícula**.

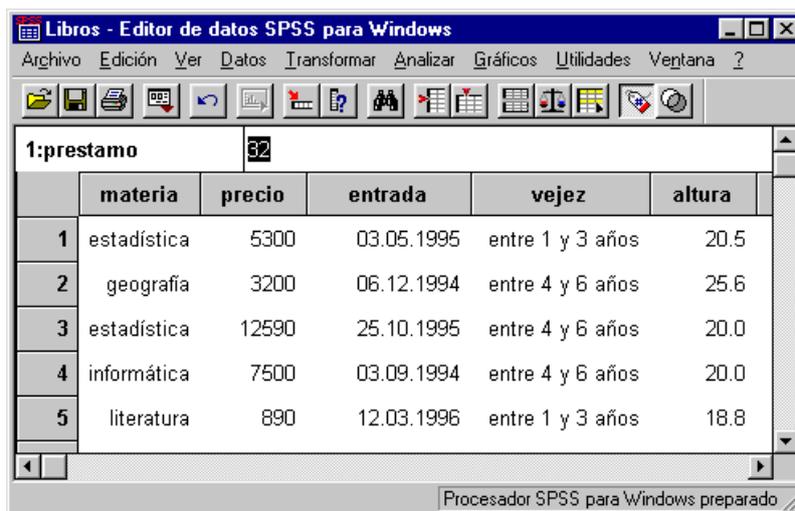


Figura 25: Editor de datos mostrando las etiquetas de las variables (y sin la cuadrícula)

Estas opciones se refieren a la apariencia del editor de datos, pero en el menú **Utilidades** nos encontramos con una posibilidad relativa al modo de operar, que es la que aparece como **Nuevo caso automático** (ver Figura 26). Por defecto aparece activada y esto origina que al traspasar los límites previos de la matriz, automáticamente se cree un nuevo caso. Si se desea anular esta manera de operar, basta con desactivar esta opción (que no esté marcada con ✓).

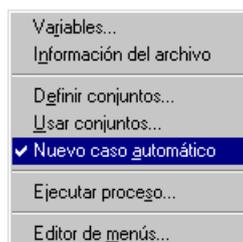


Figura 26: Menú Utilidades

()	Paréntesis
**	Exponenciación
*	Multipliación
/	División
+	Suma
-	Resta

(a) Operadores aritméticos

<	LT	Menor que
>	GT	Mayor que
<=	LE	Menor o igual que
>=	GE	Mayor o igual que
=	EQ	Igual que
≠	NE	No igual que

(b) Operadores relacionales

&	AND	Operador Y
	OR	Operador O
~	NOT	Operador NO

(c) Operadores lógicos

Cuadro 2: Operaciones aritméticas, relacionales y lógicas

## 1.12. Impresión de los datos

Para imprimir los datos hay que situarse en la ventana del editor de datos y hay que elegir **Archivo** ⇒ **Imprimir** o usar el tercer icono de la barra de herramientas. Entonces aparece un cuadro de diálogo como el de la Figura 27.



Figura 27: Cuadro de diálogo para imprimir datos

Imprimamos ahora todos los datos del fichero **Libros.sav**.

## 1.13. Importación de ficheros de datos

Acabamos de ver en los apartados anteriores los procedimientos que se emplean para generar los ficheros de datos propios de **SPSS para Windows**. En el apartado 1.10 vimos como leer (importar) este tipo de ficheros de datos. Pero en algunas ocasiones podemos disponer de datos ya grabados en un fichero generado por una aplicación informática distinta de **SPSS**. Si se desea analizar estadísticamente tal información no es preciso crear un fichero **SPSS** para tal fin, sino que basta con que se proceda a la importación del mismo. No vamos a explicar aquí detalladamente la forma de importación de estos ficheros (esto se puede consultar en la ayuda de **SPSS**), pero sí vamos a decir que es posible la importación de:

- Archivos de datos grabados en código **ASCII**, con o sin tabulación (con la opción **Archivo** ⇒ **Leer datos de texto**).
- Archivos de otros paquetes estadísticos, como **Systat** (se selecciona **Archivo** ⇒ **Abrir** y en el cuadro de diálogo que aparece se indica *Systat(\*.sys)* en *Archivos de tipo*).
- Archivos de **hojas de cálculo**, como *Lotus*, *Excel* o *formato SYLK* (se selecciona **Archivo** ⇒ **Abrir** y en *Archivos de tipo* se indica *Lotus(\*.w\*)*, *Excel(\*.xls)* o *SYLK(\*.slk)*, respectivamente).
- Archivos de **sistemas de gestión de bases de datos**, como *dBASE* (se selecciona **Archivo** ⇒ **Abrir** y en el cuadro de diálogo que aparece se indica *dBASE(\*.dbf)* en *Archivos de tipo*).
- Por supuesto, ficheros de datos **SPSS** creados en otros entornos y sistemas operativos (se selecciona **Archivo** ⇒ **Abrir** y en *Archivos de tipo* se indica *SPSS/PC+(\*.sys)* si están generados por **SPSS** en su versión para MS-DOS; o se indica *SPSS portable(\*.por)* si se trata de ficheros portátiles creados en otros entornos como Macintosh o VMS).
- Adicionalmente, y mediante ODBC, se pueden capturar y manejar datos de **Access**, **Excel** o **FoxPro** de las últimas versiones (se selecciona **Archivo** ⇒ **Captura de bases de datos** ⇒ **Nueva consulta**).

## 1.14. Transformación de variables

En este apartado vamos a ver el modo de generar nuevas variables mediante transformaciones efectuadas sobre los valores de las variables ya definidas. Cuando realicemos transformaciones en nuestro fichero **Libros.sav** grabaremos estos nuevos datos en un nuevo fichero con el nombre **Libros2.sav**

En la Tabla 2 se encuentran recogidos los operadores aritméticos, relacionales y lógicos que están permitidos. Tanto las expresiones aritméticas como las lógicas se evalúan de izquierda a derecha. Todas las expresiones entre paréntesis se evalúan antes que las que están fuera de los paréntesis y ante varios operadores en el mismo nivel, el orden de preferencia (de mayor a menor) es el que figura en la Tabla 2 (de arriba a abajo).

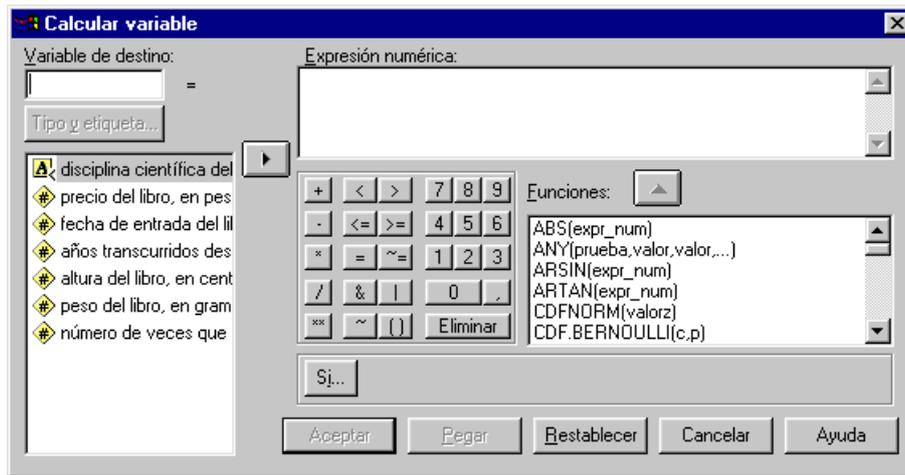


Figura 28: Cuadro de diálogo para transformar variables

Para construir una nueva variable mediante transformaciones de otras ya existentes, se tiene que elegir la opción **Transformar** ⇒ **Calcular** con lo que se abre el cuadro de diálogo de la Figura 28. En esta ventana tenemos cinco partes fundamentales: arriba a la izquierda está el lugar para escribir el nombre de la nueva variable (**Variable de destino:**), debajo aparece la lista de variables existentes, arriba a la derecha está el lugar destinado a la definición de la nueva variable (**Expresión numérica:**), debajo hay una calculadora y la lista de funciones que se pueden utilizar (**Funciones:**).

En primer lugar se asigna un nombre a la variable que queremos generar, escribiendo el mismo en el cuadro **Variable de destino**. Normalmente se va a tratar de una variable nueva, pero también cabe la posibilidad de especificar una de las ya existentes. En tal caso la modificación consistirá en sustituir los valores antiguos de la variable con los nuevos resultantes de la transformación numérica que se efectúe. Por defecto, la nueva variable será numérica. Si se quiere especificar otro tipo y añadir etiquetas, debe emplearse el método explicado en el apartado 1.5.

Una vez que se ha asignado el nombre a la variable, el siguiente paso es definir la expresión numérica que va a permitir calcular los valores de la misma. Tal expresión se escribe en el cuadro **Expresión numérica** y puede constar de los siguientes elementos: nombres de variables del fichero original, constantes, operadores y funciones. Para escribir dicha expresión ésta se puede teclear directamente pero es recomendable emplear la calculadora, la lista de variables y la lista de funciones (pulsando los correspondientes botones  y ) o haciendo doble *clic* sobre la variable o función).

Una vez que hemos terminado de escribir la expresión (utilizando, como hemos dicho, la calculadora, la lista de variables y la lista de funciones), pulsamos en **Aceptar**. Si **SPSS** encuentra un error en nuestra expresión nos lo indica convenientemente. Entonces debemos editar nuestra expresión y corregir el error, pulsando la tecla **Aceptar** de nuevo.

Inmediatamente vamos a ver ejemplos, pero no vamos a describir cada una de las funciones, pues hay muchas. Como he dicho en otros apartados, remito a las ayudas que el sistema proporciona.

#### 1.14.1. Modificación sin utilizar el condicional

En el fichero que venimos manejando figura la altura de cada libro, en centímetros, por lo que viene expresada con un decimal. Si, por ejemplo, quisiéramos expresarla en milímetros (para quitar los decimales) bastaría con multiplicar por 10. Como en este caso no es necesario generar una nueva variable, en el recuadro **Variable de destino** se escribe el nombre de la misma (**altura**). A continuación, en el recuadro **Expresión numérica** ponemos *altura\*10* (recordar que es mejor no teclearlo sino utilizar la calculadora del cuadro de diálogo y la lista de variables). Tras pulsar el botón **Aceptar**, el sistema pregunta si queremos que desaparezcan los resultados anteriores de la variable **altura** para que sean sustituidos por los nuevos valores.

Lo conveniente ahora es cambiar la definición de la variable **altura** (ver apartado 1.5) en lo que concierne al **Tipo** (*Ancho=3, Decimales=0*) y a las **Etiquetas** (*Etiqueta de variable= altura del libro, en milímetros*). Hecho esto, en la ventana del editor de datos tendríamos lo que aparece en la Figura 29. No olvidemos ahora grabar estos nuevos datos en un fichero con el nombre **Libros2.sav**.

Siguiendo con nuestro fichero de datos, en él consta una variable de tipo fecha, que es la fecha de entrada del libro a la biblioteca. Si quisiéramos saber el tiempo transcurrido entre esta fecha y la fecha actual podríamos crear una nueva variable, a la que podríamos llamar **estancia** con el resultado de la diferencia entre la fecha actual (supongamos que sea 1/2/2000) y la variable **entrada**.

Para ello, seleccionamos **Transformar** ⇒ **Calcular**. En **Variable de destino** escribimos **estancia** (ver Figura 28). Ahora tenemos que expresar la diferencia entre la fecha 1/2/2000 y las fechas de la variable **entrada**. Pero la fecha 1/2/2000 no se puede escribir directamente, sino que tenemos que utilizar una de las funciones del sistema, concretamente la función **DATE.DMY(día,mes,año)** que devuelve, en formato de fecha de **SPSS**, el valor de la fecha correspondiente al *día, mes* y *año* indicados. Por tanto, en **Expresión numérica** pondremos (con ayuda de la lista de variables y de la lista de funciones) lo siguiente: *DATE.DMY(1,2,2000)-entrada*. Después de pulsar el botón **Aceptar** obtendremos los valores que aparecen en la columna

Libros - Editor de datos SPSS para Windows

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1:altura 205

	materia	precio	entrada	vejez	altura	peso	prestar
1	E	5300	03.05.1995	1	205	325	
2	G	3200	06.12.1994	2	256	890	
3	E	12590	25.10.1995	2	200	415	
4	I	7500	03.09.1994	2	200	400	
5	L	890	12.03.1996	1	188	515	

Procesador SPSS para Windows preparado

Figura 29: Ventana del editor con la nueva variable altura

**estancia** de la Figura 30.

Como vemos, los valores son muy grandes, y esto se debe a que ambas fechas figuran en *SPSS* como el número de segundos transcurridos desde el 15 de octubre de 1582. Obviamente, la resta nos proporciona el resultado en segundos. Como estas cantidades son complicadas de manejar, vamos a realizar ahora una nueva transformación para medir dicha diferencia en días. Para realizar esto, *SPSS* dispone de una función, que es **CTIME.DAYS(valortiempo)**, que devuelve el número de días, incluyendo fracciones de día, que hay en **valortiempo**. Por tanto, generamos una nueva variable, que llamaremos **días**, que será el resultado de poner (con ayuda de la lista de variables y de la lista de funciones) en el recuadro **Expresión numérica** lo siguiente: **CTIME.DAYS(estancia)**. El resultado de esta variable se encuentra en la columna **días** de la Figura 30.

Si todavía nos parecen grandes los valores de esta nueva variable, podemos proporcionar el resultado en años. Para ello, creamos una nueva variable, que llamaremos **años**, que será el resultado de la siguiente **Expresión numérica**: **días/365.2422** (que es recomendable poner con ayuda de la calculadora del cuadro de diálogo, de la lista de variables y de la lista de funciones), pues el año tiene 365'2422 días. Los valores de esta nueva variable aparecen en la columna **años** de la Figura 30. En la línea de edición de datos se puede observar que el verdadero resultado (el almacenado por *SPSS*) tiene más de dos cifras decimales. Si queremos que en la matriz de datos aparezcan más de dos decimales tenemos que modificar el **Tipo** de la variable **años** señalando en **Decimales** el número de decimales deseado. Hagamos o no esta modificación en el *tipo* de esta variable, es conveniente colocar una etiqueta a esta nueva variable, para que sepamos en todo momento qué es lo que mide; por ejemplo, podemos escribir en **Etiqueta de variable** lo siguiente: *años de estancia del libro en la biblioteca*.

Llegado este punto voy a hacer una aclaración sobre la coma decimal que hemos tenido que escribir cuando hemos dividido por 365'2422. Es conveniente escribir la coma decimal mediante la tecla de la separación decimal que aparece en la calculadora pues aunque en ella está escrita una coma, realmente se escribe un punto al pulsar. Por tanto, si en el recuadro **Expresión numérica** del cuadro de diálogo de la Figura 28 queremos escribir un número con decimales sin emplear la calculadora, tenemos que escribir siempre un punto como separador decimal.

Aunque lo he hecho por su interés pedagógico, no es lógico conservar las variables **estancia** y **días**, por lo que podemos eliminarlas de la matriz de datos, dejando únicamente la variable **años**. Recordemos que para eliminar una variable primero se selecciona dicha variable (pulsando en su nombre) y luego se elige **Edición** ⇒ **Eliminar** o se pulsa el botón auxiliar del ratón y del menú contextual que aparece se elige **Eliminar**.

Otra opción para generar directamente la variable **años**, sin tener que generar las variables **estancia** y **días** es poner directamente (con ayuda de la calculadora del cuadro de diálogo, de la lista de variables y de la lista de funciones) en **Expresión numérica** todas las operaciones que hemos realizado: **(CTIME.DAYS( DATE.DMY(1,2,2000)-entrada))/365.2422**

Por último, si queremos que la nueva variable **años** aparezca en el editor de datos junto a la variable **entrada**, a su derecha (ver Figura 31) insertamos primero una variable vacía (que automáticamente se llamará **var0001**) entre las variables **entrada** y **vejez**; luego seleccionamos la variable **años**, la *cortamos*, seleccionamos **var0001** y *pegamos* (para hacer este proceso consultar el apartado 1.11).

Ya que hemos empleado algunas de las funciones de fecha y hora, voy a listar, a continuación, las funciones de fecha y hora disponibles en *SPSS*:

**CTIME.DAYS(valortiempo)** Su resultado es numérico. Devuelve el número de días, incluyendo fracciones de día, que hay en **valortiempo**, el cual debe ser un número o una expresión en formato de tiempo.

**CTIME.HOURS(valortiempo)** Su resultado es numérico. Devuelve el número de horas, incluyendo las fracciones de hora, en **valortiempo**, el cual debe ser un número o una expresión con formato de tiempo.

**CTIME.MINUTES(valortiempo)** Su resultado es numérico. Devuelve el número de minutos, incluyendo las fracciones de minuto, en **valortiempo**, el cual debe ser un número o una expresión con formato de tiempo.

	peso	prestamo	estancia	días	años	var
1	325	32	1,50E+08	1735,00	4,75	
2	890	18	1,63E+08	1883,00	5,16	
3	415	9	1,35E+08	1560,00	4,27	
4	400	15	1,71E+08	1977,00	5,41	
5	515	2	1,23E+08	1421,00	3,89	
6	650	36	1,14E+08	1315,00	3,60	
7	790	5	1,36E+08	1570,00	4,30	
8	890	9	1,64E+08	1893,00	5,18	
9	320	7	1,76E+08	2035,00	5,57	
10	420	12	1,50E+08	1734,00	4,75	

Figura 30: Ventana del editor con las nuevas variables estancia, días y años

	materia	precio	entrada	años	vejez	altura	peso
1	E	5300	03.05.1995	4,75	1	205	32
2	G	3200	06.12.1994	5,16	2	256	89
3	E	12590	25.10.1995	4,27	2	200	41
4	I	7500	03.09.1994	5,41	2	200	40
5	L	890	12.03.1996	3,89	1	188	51

Figura 31: Ventana del editor con la variable años junto a la variable entrada

**CTIME.SECONDS(valortiempo)** Su resultado es numérico. Devuelve el número de segundos, incluyendo las fracciones de segundo, en **valortiempo**, el cual debe ser un número o una expresión con formato de tiempo.

**DATE.DMY(día,mes,año)** Su resultado es numérico en el formato de fecha. Devuelve un valor de fecha correspondiente al **día**, **mes** y **año** indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. Los argumentos deben ser enteros. El **día** debe estar entre 1 y 31, el **mes** entre 1 y 12, y, por último, el **año** debe ser un entero de cuatro dígitos mayor que 1582.

**DATE.MDY(mes,día,año)** Su resultado es numérico en el formato de fecha. Devuelve un valor de fecha correspondiente al **mes**, **día** y **año** indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. Los argumentos son iguales a los de la función anterior.

**DATE.MOYR(mes,año)** Su resultado es numérico en el formato de fecha. Devuelve un valor de fecha correspondiente al **mes** y **año** indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. Los argumentos deben ser enteros. El **mes** debe estar entre 1 y 12, y el **año** debe ser un entero de cuatro dígitos mayor que 1582.

**DATE.QYR(trimestre,año)** Su resultado es numérico en el formato de fecha. Devuelve un valor de fecha correspondiente al **trimestre** y **año** indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. Los argumentos deben ser enteros. El **trimestre** debe estar entre 1 y 4, y el **año** debe ser un entero con cuatro dígitos mayor que 1582.

**DATE.WKYR(núm.semana,año)** Su resultado es numérico en el formato de fecha. Devuelve un valor de fecha correspondiente al

**núm\_semana** y **año** indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. El parámetro **núm\_semana** debe estar entre 1 y 52, y el **año** debe ser un entero de cuatro dígitos mayor que 1582.

**DATE.YRDAY(año,núm.día)** Su resultado es numérico en el formato de fecha. Devuelve un valor de fecha correspondiente al **núm.día** y **año** indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. Los argumentos deben ser enteros. El parámetro **núm.día** debe estar entre 1 y 366, y el **año** debe ser un entero de cuatro dígitos mayor que 1582.

**TIME.DAYS(días)** Su resultado es numérico en el formato de intervalo de tiempo. Devuelve un intervalo de tiempo correspondiente al número de días indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. El argumento debe ser numérico.

**TIME.HMS(horas,min,seg)** Su resultado es numérico en el formato de intervalo de tiempo. Devuelve un intervalo de tiempo correspondiente al número de horas, minutos y segundos indicados. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*. Los argumentos deben ser enteros no mayores que 24, 60 y 60, respectivamente, con la excepción del primer argumento no nulo, que puede exceder su límite, y del último argumento, que puede tener una parte fraccional.

**XDATE.DATE(valorfecha)** Su resultado es numérico en el formato de fecha. Devuelve la porción de fecha de un valor numérico en el formato de fecha. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha*.

**XDATE.HOUR(valorfecha)** Su resultado es numérico. Devuelve la hora (un entero entre 0 y 23) a partir de un valor numérico en el formato de fecha.

**XDATE.JDAY(valorfecha)** Su resultado es numérico. Devuelve el día del año (un entero entre 1 y 366) a partir de un valor numérico en el formato de fecha.

**XDATE.MDAY(valorfecha)** Su resultado es numérico. Devuelve el día del mes (un entero entre 1 y 31) a partir de un valor numérico en el formato de fecha.

**XDATE.MINUTE(valorfecha)** Su resultado es numérico. Devuelve el minuto (un entero entre 0 y 59) a partir de un valor numérico en el formato de fecha.

**XDATE.MONTH(valorfecha)** Su resultado es numérico. Devuelve el mes (un entero entre 1 y 12) a partir de un valor numérico en el formato de fecha.

**XDATE.QUARTER(valorfecha)** Su resultado es numérico. Devuelve el trimestre del año (un entero entre 1 y 4) a partir de un valor numérico en el formato de fecha.

**XDATE.SECOND(valorfecha)** Su resultado es numérico. Devuelve el segundo (un entero entre 0 y 60) a partir de un valor numérico en el formato de fecha.

**XDATE.TDAY(valortiempo)** Su resultado es numérico. Devuelve el número de días completos (como un entero) a partir de un valor numérico en el formato de intervalo de tiempo

**XDATE.TIME(valorfecha)** Su resultado es numérico en el formato de intervalo de tiempo, que representa el número de segundos desde la medianoche. Devuelve el momento del día a partir de un valor numérico en el formato de fecha. Para mostrar este valor correctamente, hay que definir esta nueva variable con un formato de tipo *Fecha* (horas y minutos).

**XDATE.WEEK(valorfecha)** Su resultado es numérico. Devuelve el número de la semana (un entero entre 1 y 52) a partir de un valor numérico en el formato de fecha.

**XDATE.WKDAY(valorfecha)** Su resultado es numérico. Devuelve el número del día de la semana (un entero entre 1 -Domingo- y 7 -Sábado-) a partir de un valor numérico en el formato de fecha.

**XDATE.YEAR(valorfecha)** Su resultado es numérico. Devuelve el año (un entero de cuatro dígitos) a partir de un valor numérico en el formato de fecha.

**YRMODA(año,mes,día)** Su resultado es numérico. Devuelve el número de días desde el 15 de octubre de 1582 hasta la fecha representada por los argumentos **año**, **mes** y **día**, los cuales deben ser enteros que compongan una fecha válida desde el 15 de octubre de 1582. Para los valores de dos dígitos se calcula el año completo a partir de lo especificado en las opciones de datos (menú **Edición** ⇒ **Opciones** ⇒ *Pestaña de Datos*).

### 1.14.2. Modificación condicional de variables

En el apartado anterior hemos descrito las posibilidades que brinda el cuadro de diálogo correspondiente a la opción **Transformar** ⇒ **Calcular**; pero hemos hecho caso omiso a un botón: el de **Si...** condicional, que sirve para activar la modificación condicional de las variables. Tal posibilidad debe emplearse cuando se quiere que las modificaciones especificadas en la expresión numérica no afecten a todos los casos, sino sólo a aquellos que reúnan una serie de características. Tales características se especifican mediante expresiones condicionales, de forma que los nuevos valores sólo van a ser calculados para los individuos que satisfagan una serie de condiciones lógicas.

Par tal fin, una vez seleccionada la opción **Transformar** ⇒ **Calcular**, debe pulsarse el botón **Si...**, con lo que se abre el cuadro de la Figura 32. Este cuadro tiene una apariencia similar al inicial, con la diferencia de que el cuadro de texto sirve para especificar las expresiones condicionales. El uso de dicho cuadro y las normas de sintaxis de este tipo de expresiones son las mismas que para las numéricas, por lo que me voy a limitar a exponer sucintamente algunos ejemplos.

Supongamos que, en el ejemplo que venimos manejando, quisiéramos calcular el producto del **peso** por la **altura** sólo para los libros de *estadística*. Entonces, en el cuadro de diálogo de la Figura 32, hay que activar la opción **Incluir si el caso satisface**

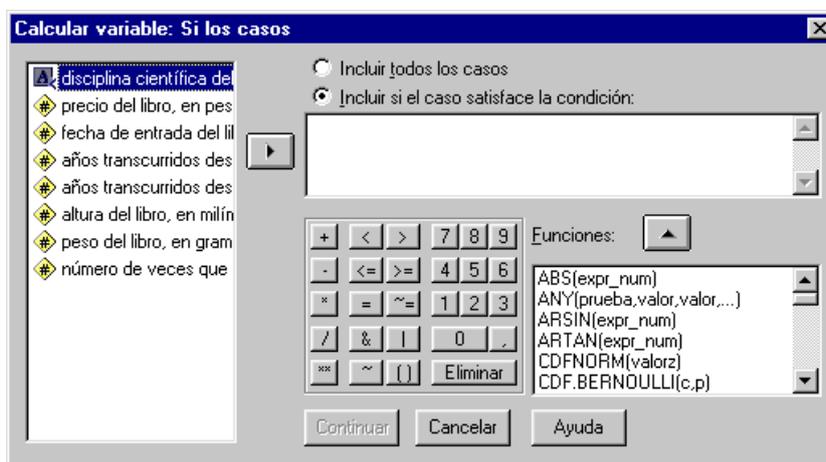


Figura 32: Cuadro de diálogo para hacer la modificación condicional de una variable

la **condición**. A continuación hay que poner (como siempre, con ayuda de la lista de variables) en el recuadro en blanco la expresión: *materia='E'*. Observemos que el dato *E* va entre dos comillas *simples*. Poner estas comillas es obligatorio siempre que el resultado sea una cadena de caracteres. (En el teclado del ordenador esta comilla simple está situada en la primera fila superior de teclas, sin contar la fila de funciones F1, F2, etc., y habitualmente está en la misma tecla que el signo de interrogación.) Después se pulsa el botón **Continuar** y volvemos al cuadro de diálogo de la Figura 28. En este cuadro de diálogo, junto a **Variable de destino** escribimos, por ejemplo, **robustez**; y debajo de **Expresión numérica** ponemos *peso\*altura* (como ya he dicho varias veces, con ayuda de la calculadora de este cuadro de diálogo y de la lista de funciones). Tras pulsar **Aceptar** obtenemos en el editor de datos una nueva columna con la nueva variable calculada sólo para los casos en los que la variable **materia** es igual a *estadística* (ver Figura 33). El resto de los casos están en blanco (son valores perdidos o ausentes). Ya que la nueva variable **robustez** está relacionada con la *altura* y con el *peso*, ahora podríamos cambiar dicha

	años	vejez	altura	peso	prestamo	robustez
11	4,32	3	235	620	42	145700
12	3,14	2	235	720	20	169200
13	4,73	3	309	850	10	.
14	5,22	2	205	620	9	.
15	3,96	1	220	475	15	104500

Figura 33: Ventana del editor con la nueva variable robustez

variable de posición colocándola junto a la variable peso, a su derecha. Sería conveniente ahora volver a grabar el fichero de datos **Libros2.sav**.

Supongamos ahora que una determinada transformación queremos que se realice sólo para los libros de estadística cuyo precio sea menor que 10000 pesetas; entonces la expresión condicional que hay que poner en el recuadro en blanco del cuadro de diálogo de la Figura 32 sería (ver la Tabla 2): *materia='E' AND precio LT 10000* o también: *materia='E' & precio<10000* (el uso de los espacios en blanco es optativo).

### 1.14.3. Recodificación de variables

Después de haber visto cómo es posible generar nuevas variables mediante transformaciones numéricas (condicionales o no), vamos ahora a discutir la manera de cambiar el plan de codificación de una variable determinada.

La recodificación de variables puede ser necesaria por varios motivos: porque se ha visto que el plan de codificación inicial es erróneo, porque ha habido una ampliación en la muestra y por consiguiente han aparecido nuevos valores que no estaban inicialmente contemplados, porque puede facilitar la realización de determinados análisis estadísticos posteriores,

porque se quiere combinar categorías y definir intervalos, etc.

Antes de pasar a un ejemplo, digamos que la recodificación se puede llevar a cabo de dos maneras: 1) se puede recodificar guardando el nuevo plan de codificación en la misma variable (con lo que se perderán los antiguos valores), o 2) se puede mantener la variable original en su estado inicial y generar una nueva variable cuyos valores van a ser el resultado de aplicar el nuevo plan de codificación a los de la antigua (de esta manera el fichero va a contar con dos variables que representan el mismo fenómeno pero con dos planes de codificación distintos). Nosotros vamos a ver un ejemplo de este segundo caso.

Supongamos que en nuestro fichero **Libros2.sav** queremos definir unos intervalos para que la variable **préstamo** sea una variable ordinal y podamos, posteriormente, ver si existe relación entre ella y la otra variable ordinal: **vejez**. Por ejemplo, podemos hacer la siguiente asignación:

valor nuevo	valor antiguo
1	desde el menor hasta 10
2	entre 11 y 20
3	entre 21 y 30
4	entre 31 y el mayor valor

Los pasos a seguir son los siguientes: primero seleccionamos la opción **Transformar** ⇒ **Recodificar** ⇒ **En distintas variables**. Entonces aparece el cuadro de diálogo de la Figura 34. En este cuadro de diálogo seleccionamos la variable que queremos



Figura 34: Cuadro de diálogo para recodificar una variable

recodificar (haciendo doble *click* sobre ella en la lista de variables que hay a la izquierda o haciendo un *click* y luego pulsando el botón ). En **Variable de resultado** escribimos **pres\_cat** debajo de **Nombre** y **Categorías de la variable préstamo** en **Etiqueta**. Después pulsamos el botón **Cambiar** y a continuación el botón **Valores antiguos y nuevos**, con lo que nos aparece el cuadro de diálogo de la Figura 35. En este cuadro tenemos que hacer la asignación deseada de forma muy similar a como hicimos la

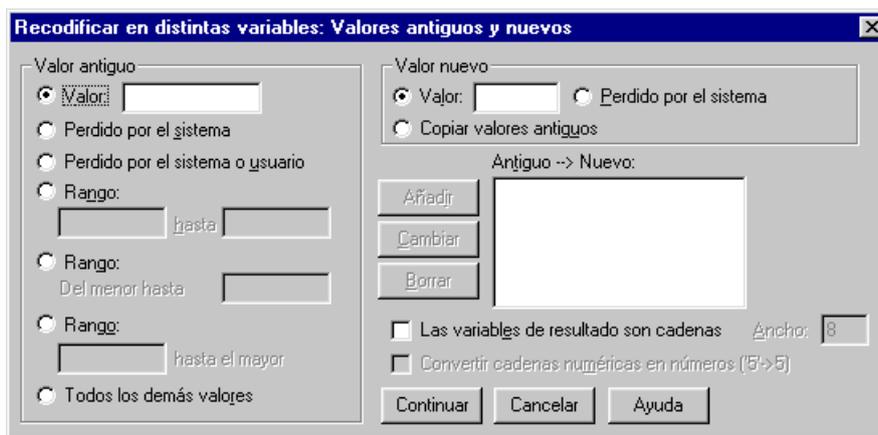


Figura 35: Cuadro de diálogo para hacer la asignación de nuevos códigos

asignación de etiquetas de una variable:

- En **Valor nuevo**, junto a **Valor** escribimos 1. En **Valor antiguo** seleccionamos **Rango: Del menor hasta** y en éste escribimos 10 (esto significa desde el menor valor de la variable hasta 10). Luego pulsamos el botón **Añadir**, con lo que este cambio queda reflejado en el cuadro que hay bajo **Antiguo → Nuevo**.
- En **Valor nuevo**, junto a **Valor** escribimos 2. En **Valor antiguo** seleccionamos **Rango:** y escribimos 11 **hasta** 20. A continuación pulsamos el botón **Añadir**.

- En **Valor nuevo**, junto a **Valor** escribimos 3. En **Valor antiguo** seleccionamos **Rango:** y escribimos 21 hasta 30. A continuación pulsamos el botón **Añadir**.
- En **Valor nuevo**, junto a **Valor** escribimos 4. En **Valor antiguo** seleccionamos **Rango: hasta el mayor** y escribimos 31 (esto quiere decir desde 31 hasta el mayor) . A continuación pulsamos el botón **Añadir**.

Una vez concluida la asignación, pulsamos **Continuar**, con lo que el sistema nos devuelve al cuadro de la Figura 34. En este cuadro pulsamos el botón **Aceptar**. Como consecuencia, en el editor de datos tenemos una columna más con la variable **pres\_cat**. Como esta variable sólo toma los valores 1, 2, 3 y 4, debemos cambiar un poco su definición. Concretamente, en **Tipo** debemos escribir un 1 en **Ancho** y un 0 en **Decimales**. Y en **Medida** debemos seleccionar **Ordinal**. El editor de datos aparecería como en la Figura 36.

	vejez	altura	peso	robustez	prestamo	pres_cat
1	1	205	325	66625	32	4
2	2	256	890	.	18	2
3	2	200	415	83000	9	1
4	2	200	400	.	15	2
5	1	188	515	.	2	1

Figura 36: Ventana del editor con la nueva variable pres\_cat (categorías de la variable préstamo)

#### 1.14.4. Recodificación automática de variables

Este tipo de modificación consiste en la ordenación ascendente o descendente del conjunto de valores de una variable. Es aplicable tanto a variables numéricas como no numéricas. En estas últimas el criterio es el orden alfabético de los caracteres donde las mayúsculas preceden a las minúsculas.

Este tipo de modificación se realiza tras seleccionar **Transformar** ⇒ **Recodificación automática**. Al hacerlo se abre el cuadro de diálogo de la Figura 37.

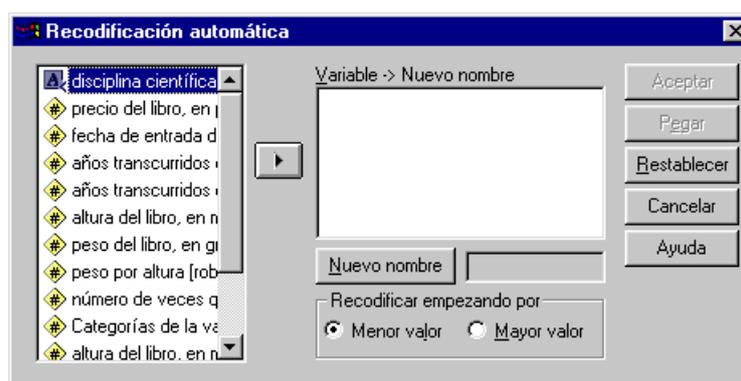


Figura 37: Cuadro de diálogo para la recodificación automática

Supongamos, por ejemplo, que queremos ordenar los resultados de la variable **préstamo** en orden ascendente. En el cuadro de diálogo de la Figura 37 ponemos **préstamo** en el recuadro **Variable** → **Nuevo nombre** (haciendo doble *click* en el nombre de la variable o haciendo un *click* en ella y luego pulsando el botón ). A continuación escribimos junto a **Nuevo nombre** el nombre deseado, por ejemplo **ord\_pres**, y pulsamos en **Nuevo nombre**. Después seleccionamos **Recodificar empezando por, Menor valor**. Por último pulsamos **Aceptar**. El resultado es la aparición en el editor de datos de la nueva variable **ord\_pres** (Figura 38). Como la etiqueta de la nueva variable **ord\_pres** es la misma que la de la variable **préstamo**, es conveniente ahora cambiarla; por ejemplo, podemos poner *Ordenación de la variable préstamo* en **Etiqueta de variable**.

	vejez	altura	peso	prestamo	ord_pres	var
15	1	22.0	475	15	12	
16	1	20.0	320	1	1	
17	1	21.0	380	15	12	
18	2	35.9	790	16	13	
19	1	33.0	780	9	7	
20	2	20.0	500	16	13	
21	2	25.9	550	11	9	
22	1	25.0	820	8	6	
23	1	19.8	360	4	3	
24	4	25.3	490	4	3	
25	4	22.8	480	2	2	

Figura 38: Editor de datos con la nueva variable ord\_pres, generada por recodificación automática de la variable préstamo

#### 1.14.5. Ordenación por rangos

**Observación** Este apartado se entenderá mejor cuando se haya estudiado la relación entre dos variables ordinales, pues en este tema aparece el concepto de rango.

En el apartado anterior hemos ordenado de menor a mayor (o de mayor a menor) los resultados de una variable. Ahora ordenaremos por rangos los casos de la variable original.

Para hacer la ordenación por rangos hay que seleccionar la opción **Transformar** ⇒ **Asignar rangos a los casos** y aparecerá el cuadro de diálogo de la Figura 39. En este cuadro, lo primero es seleccionar la variable cuyos valores se van a ordenar. En este

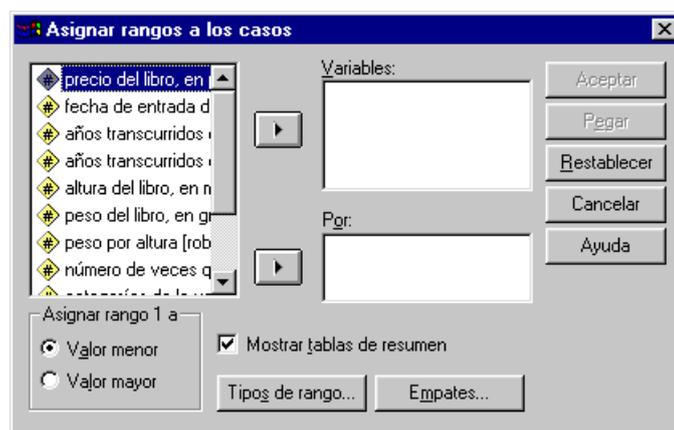


Figura 39: Cuadro de diálogo para asignar rangos a los casos de una variable

caso, aunque siempre se genera una nueva variable, no hace falta especificar su nombre, pues éste va a ser automáticamente asignado por **SPSS** anteponiendo una **r** al original. Igualmente, el sistema va a dotar a la nueva variable de una etiqueta en la que se especifica el método de ordenación adoptado.

Si lo que se quiere es una ordenación simple según el criterio ascendente, una vez seleccionada la variable original basta con pulsar el botón **Aceptar**. Por ejemplo, la asignación de rangos a la variable del apartado anterior (**préstamo**) daría lugar a una nueva variable cuyos valores se puede ver en la Figura 40. Como siempre, podemos cambiar el nombre de la variable **rpréstamo** o cualquier otra especificación de ella. Por ejemplo, recomendaría cambiar la **Etiqueta de variable: RANK of PRESTAMO** por *Rangos de la variable préstamo*. También es conveniente cambiar el **Tipo** escribiendo 4 en **Ancho** y 1 en **Decimales**. En **Medida** deberíamos señalar **Ordinal**. La nueva variable aparece en la Figura 40. Como vemos, es una ordenación

ascendente por rangos simples (el valor más bajo de la variable original es el primero de la nueva), donde los empates han sido resueltos hallando la media aritmética de los rangos consecutivos que tendrían si no estuviesen empatados.

	altura	peso	prestamo	ord_pres	rprestam	va
15	22.0	475	15	12	19,0	
16	20.0	320	1	1	1,0	
17	21.0	380	15	12	19,0	
18	35.9	790	16	13	22,0	
19	33.0	780	9	7	11,0	
20	20.0	500	16	13	22,0	
21	25.9	550	11	9	15,0	
22	25.0	820	8	6	8,0	
23	19.8	360	4	3	4,5	
24	25.3	490	4	3	4,5	
25	22.8	480	2	2	2,5	

Figura 40: Editor de datos con la nueva variable rprestam, generada mediante asignación de rangos a los casos de la variable préstamo

Este es el procedimiento establecido por defecto. Si queremos realizar otro tipo de ordenaciones, se debe recurrir al empleo del resto de los componentes del cuadro de la Figura 39. Por ejemplo, si pulsamos el botón **Empates** y en el cuadro de diálogo que aparece seleccionamos **Rangos secuenciales para valores únicos** entonces el resultado es una variable exactamente igual a la obtenida mediante recodificación automática.

### 1.14.6. Otras modificaciones

La opción **Categorizar variables** del menú **Transformar** convierte datos numéricos continuos en un número discreto de categorías. Este procedimiento crea nuevas variables que contienen los datos categóricos. Los datos se categorizan según grupos de percentiles; cada grupo contiene aproximadamente el mismo número de casos. Por ejemplo, una especificación de cuatro grupos asignaría un valor de 1 a los casos situados bajo el percentil al 25 % (primer cuartil), 2 a los casos entre el percentil al 25 % y el percentil al 50 % (segundo cuartil o mediana), 3 a los casos situados entre el percentil al 50 % y el percentil al 75 % (tercer cuartil) y 4 a los casos por encima del percentil 75 %.

La opción **Contar apariciones** del menú **Transformar** permite generar una nueva variable cuyos valores vienen dados, para cada caso, por el número de veces que se da un valor o conjunto de valores determinado en una lista de variables del fichero de trabajo original. En definitiva, el sistema cuenta las ocurrencias de un valor en una o varias variables y el resultado pasa a ser el valor de la nueva variable.

La opción **Reemplazar valores perdidos** del menú **Transformar** permite generar nuevas variables en las que los valores perdidos o ausentes que hubiese en las originales serán sustituidos por estimaciones (esto se entenderá mejor cuando se haya estudiado la parte de Estimación de Parámetros).

También he de decir que normalmente todas las transformaciones que hemos visto son procesadas de forma inmediata una vez que el correspondiente botón es pulsado; pero es posible optar por retrasar la ejecución de las mismas. Esto puede ser conveniente cuando el fichero de trabajo es muy grande y hay muchas transformaciones que realizar sobre el mismo. Esto implica cambiar uno de los parámetros que el sistema tiene establecidos por defecto, y para ello debe utilizarse la opción **Edición ⇒ Opciones ⇒ Pestaña de Datos** (ver Figura 11).

### 1.15. Ponderación de los casos

La opción **Datos ⇒ Ponderar casos** proporciona a los casos diferentes ponderaciones (mediante una réplica simulada) para el análisis estadístico.

- Los valores de la variable de ponderación deben indicar el número de observaciones representadas por casos únicos en el archivo de datos.
- Los casos con valores perdidos, negativos o cero para la variable de ponderación se excluyen del análisis.
- Los valores fraccionarios son válidos; se usan exactamente donde adquieren sentido y, con mayor probabilidad, donde se tabulan los casos.

Si se aplica una variable de ponderación, ésta seguirá vigente hasta que se seleccione otra o se desactive la ponderación. Si se guarda un archivo de datos ponderado, la información de ponderación se guardará con el archivo. Se puede desactivar la ponderación en cualquier momento, incluso después de haber guardado el archivo de forma ponderada.

- ◇ Ponderaciones en las tablas de contingencia: En los procedimientos de tablas de contingencia, las frecuencias de casilla basadas en ponderaciones fraccionales se redondean hasta el valor entero más próximo. Por ejemplo, una frecuencia de casilla de 4.2 basada en una ponderación fraccional se redondea a 4.
- ◇ Ponderaciones en los diagramas de dispersión y los histogramas: Los diagramas de dispersión y los histogramas tienen una opción para activar y desactivar las ponderaciones de los casos, pero dicha opción no afecta a los casos que tienen un valor negativo, un valor 0 o un valor perdido para la variable de ponderación. Estos casos permanecen excluidos del gráfico incluso si se desactiva la ponderación desde el gráfico.

Vamos ahora a crear un nuevo fichero de datos que tenga en la primera columna los resultados de una variable aleatoria discreta y en la segunda columna sus correspondientes frecuencias absolutas. Por ejemplo, vamos a introducir los datos del número de palabras por línea de una página de un libro:

Nº de palabras	4	5	8	9	10	11	12	13	14	15	16	17
Nº de líneas	1	1	2	3	2	7	11	14	3	2	1	1

El número de palabras por línea es la variable y el número de líneas es la frecuencia absoluta. Para introducir estos datos hacemos lo siguiente:

- ♡ Seleccionamos la opción **Archivo** ⇒ **Nuevo** ⇒ **Datos**.
- ♡ Definimos la columna primera (que es realmente la única variable) con las siguientes especificaciones: *Nombre de variable:* **palabras**; *Tipo:* Numérico, *Ancho=* 2, *Decimales=* 0; *Etiqueta de variable:* Número de palabras por línea de texto; *Formato de columna:* Ancho=8; *Medida:* Escala
- ♡ Definimos la columna segunda (que son las frecuencias) con las siguientes especificaciones: *Nombre de variable:* **frecuenc**; *Tipo:* Numérico, *Ancho=* 2, *Decimales=* 0; *Etiqueta de variable:* Frecuencia absoluta de la variable palabras; *Formato de columna:* Ancho=8; *Medida:* Escala
- ♡ Introducimos los datos de la forma en que ya sabemos.
- ♡ Seleccionamos la opción **Datos** ⇒ **Ponderar casos**. Nos aparece el cuadro de diálogo de la Figura 41 En este cuadro de

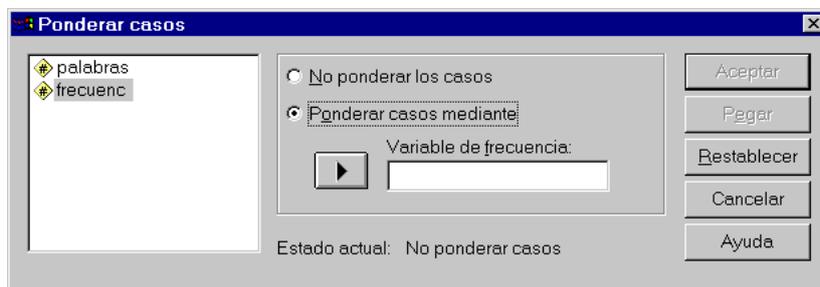


Figura 41: Cuadro de diálogo para ponderar los casos de una variable

diálogo se activa la opción **Ponderar casos mediante**, y en **Variable de frecuencia** se selecciona (mediante la lista de columnas que hay a la izquierda) la columna **frecuenc**. A continuación se pulsa en **Aceptar** y así la ponderación queda efectiva.

- ♡ Grabamos este nuevo fichero de datos con el nombre **PalabrasPorLinea.sav**. Así, la apariencia del editor de datos será la de la Figura 42.

Este fichero de datos lo utilizaremos en la sección 2.3 para calcular las medidas descriptivas de la variable *Número de palabras por línea*.

Figura 42: Ventana del editor de datos con los resultados de la variable palabras

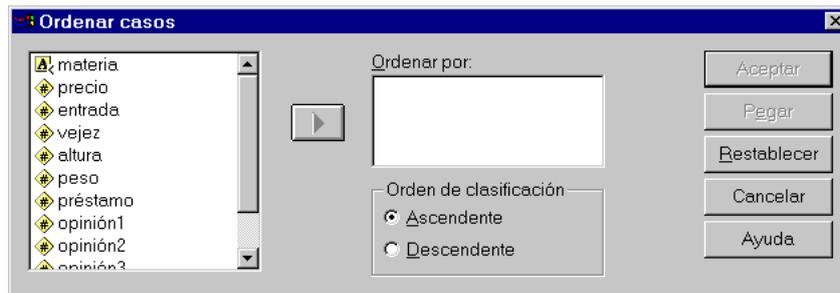


Figura 43: Cuadro de diálogo para hacer una ordenación de los casos

### 1.16. Ordenación de los casos

La opción **Datos** ⇒ **Ordenar casos** ordena los casos (las filas) del archivo de datos basándose en los valores de una o más variables de ordenación. Se puede ordenar los casos en orden ascendente o descendente. El cuadro de diálogo que aparece es

Figura 44: Editor de datos con la reordenación de los casos según la variable precio

el de la Figura 43.

- Si se selecciona más de una variable de ordenación, los casos se ordenarán por variable dentro de las categorías de la variable anterior de la lista **Ordenar por:**.
- Para las variables de cadena, las letras mayúsculas preceden a las minúsculas correspondientes en orden. Por ejemplo, el valor de cadena “No” precede a “no”.

Si de nuestro fichero de datos **Libros2.sav** seleccionamos la variable **precio** para ponerla en el recuadro **Ordenar por:** entonces se produce la reordenación de los 30 casos que aparece en la Figura 44.

## 2. Estadística descriptiva unidimensional

### 2.1. Distribución de frecuencias

Para obtener la distribución de frecuencias, las medidas descriptivas de los datos y las representaciones gráficas, se selecciona la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**. Entonces aparece el cuadro de diálogo de la Figura 45. En

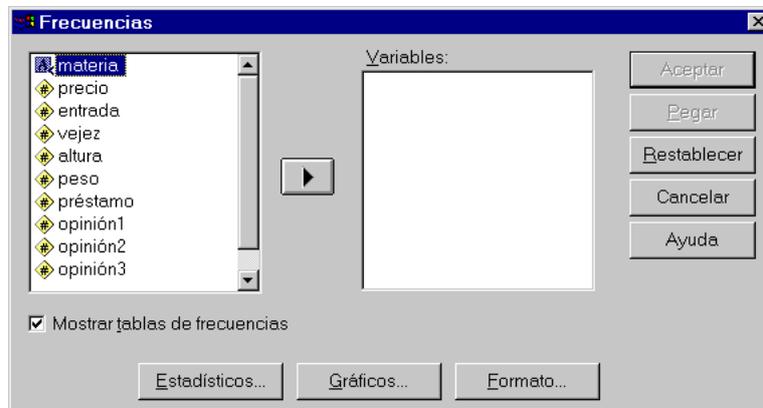


Figura 45: Cuadro de diálogo de la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**

este cuadro se selecciona, en primer lugar, la variable o variables objeto del procedimiento. Para ello se marcan las mismas en la lista de variables y se pasan al recuadro de la derecha pulsando el botón . También es posible seleccionar y pasar directamente haciendo doble *clic* sobre los nombres de dichas variables. Como podemos ver, la opción **Mostrar tablas de frecuencias** está activada. Esto quiere decir que si sólo queremos obtener la distribución de frecuencias basta con pulsar **Aceptar**. Por ejemplo, si del fichero de datos **Libros.sav** seleccionamos la variable **materia** y luego pulsamos el botón **Aceptar** entonces aparece la *ventana del visor de resultados* de la Figura 46.

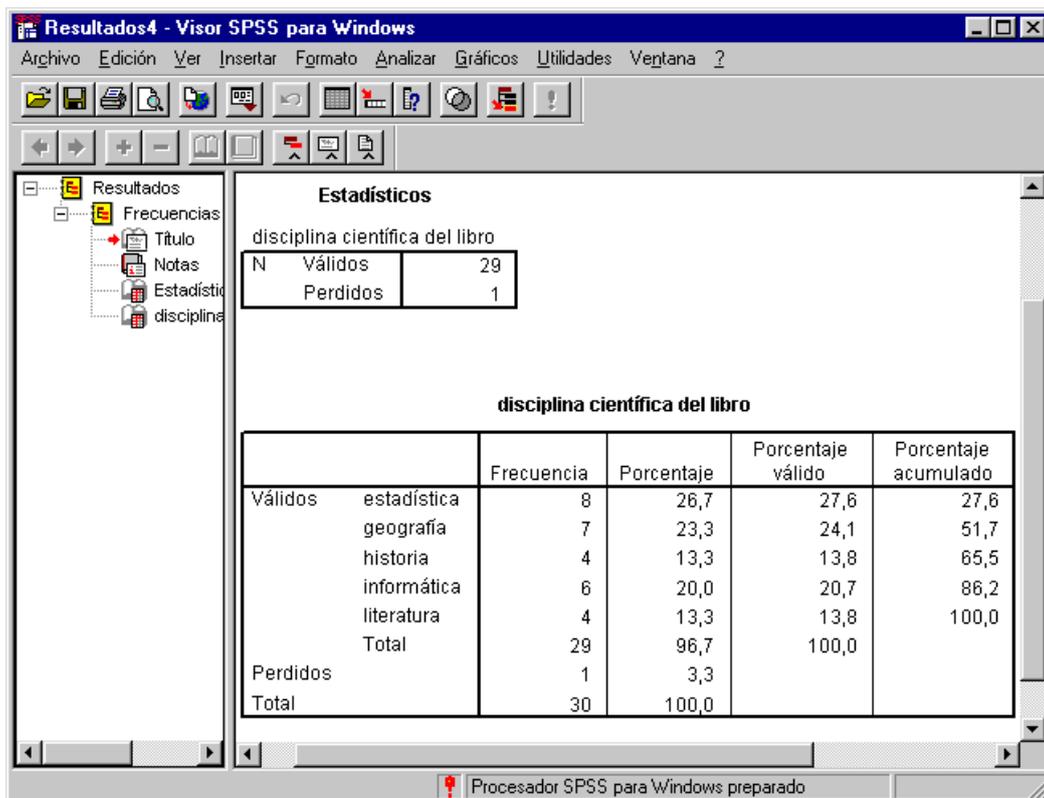


Figura 46: Visor con la distribución de frecuencias de la variable **materia** (*disciplina científica del libro*)

## 2.2. Observaciones sobre la presentación, modificación y grabación de los resultados en la ventana del visor de SPSS

### 2.2.1. Cómo controlar la visualización de la información sobre variables y valores de datos en las tablas pivote y en los titulares

Ahora que hemos obtenido las dos primeras tablas pivote (las dos tablas que se pueden ver en la *ventana del visor de resultados* de la Figura 46) quisiera señalar que es posible controlar la visualización de la información sobre variables y valores de datos en las tablas pivote y en los titulares. Esta tarea se realiza mediante **Edición** ⇒ **Opciones** ⇒ **Pestaña de Etiquetas de los resultados**. Entonces aparece el cuadro de diálogo de la Figura 47.

Se pueden visualizar los nombres de las variables, las etiquetas de las variables, los valores reales de los datos, las etiquetas definidas para estos datos, o una combinación de estas opciones. Tengamos en cuenta que las etiquetas de las variables y las etiquetas de los valores pueden facilitar la interpretación de los resultados; pero también es cierto que las etiquetas largas pueden crear dificultades en algunas tablas. Por tanto, en cada caso elegiremos las opciones que más nos convengan.

Las opciones de etiquetas de los resultados afectan sólo a los nuevos resultados que se producen después de cambiar las selecciones. A los resultados que ya se muestran en el visor no les afectan los cambios en la configuración. Esta configuración sólo afecta a los resultados de las tablas pivote. Los resultados de texto no se verán afectados por estas selecciones.

Por ejemplo, si queremos que tanto en los títulos como dentro de las tablas pivote figuren sólo los nombres (y no las etiquetas) de las variables y de sus resultados, tenemos que elegir las cuatro opciones que están seleccionadas en la Figura 47.

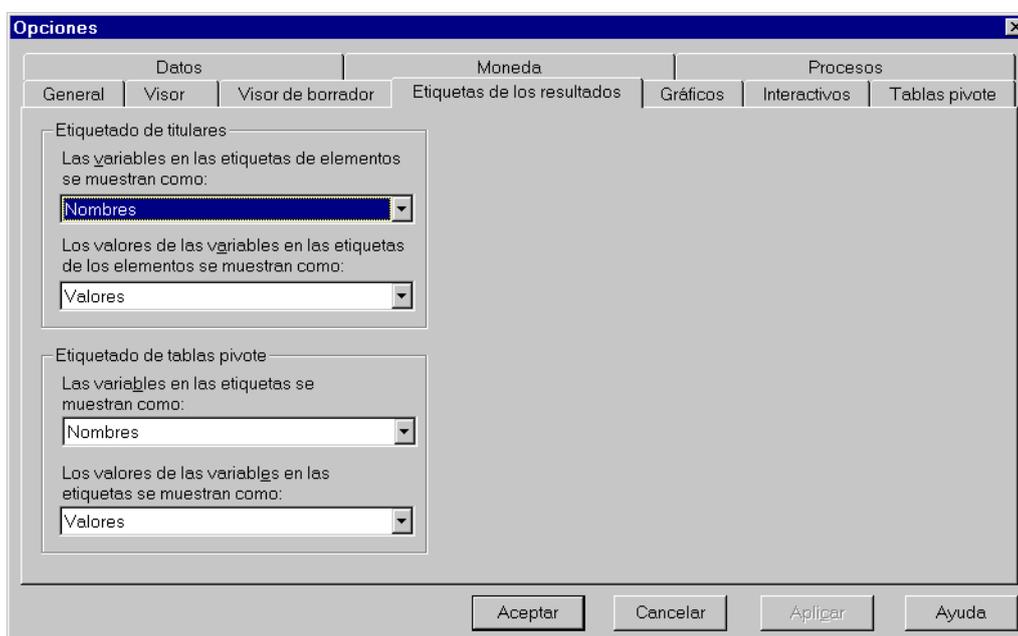


Figura 47: Pestaña de Etiquetas de los resultados, dentro de la opción Edición ⇒ Opciones

### 2.2.2. Cómo establecer el aspecto de las tablas pivote por defecto

Para establecer el aspecto de tabla por defecto, utilizado para los resultados de nuevas tablas pivote, de la barra de menú hay que seleccionar **Edición** ⇒ **Opciones** ⇒ **Pestaña de Tablas pivote**. Aparece entonces el cuadro de diálogo de la Figura 48.

El apartado **Aspectos de tabla** puede controlar una variedad de atributos de tabla pivote, incluyendo la visualización y el ancho de las líneas de cuadrícula; el estilo, tamaño y color de la fuente; así como los colores de fondo. Se selecciona un aspecto de tabla de la lista de archivos y se pulsa en **Aceptar** o **Aplicar**. Por defecto, la lista muestra los aspectos de tabla guardados en el directorio LOOKS del directorio donde se instaló el programa. Se puede utilizar uno de los aspectos de tabla que se incluyen en *SPSS* o bien crear uno propio en el editor de tablas pivote, para lo cual se hace doble *clic* sobre la tabla pivote (para poder editarla) y se selecciona **Formato** ⇒ **Aspectos de tabla**.

El botón **Examinar** permite seleccionar un aspecto de tabla de otro directorio.

El botón **Establecer directorio de aspectos** permite cambiar el directorio de aspectos por defecto.

El apartado **Ajustar los anchos de columna para** controla el ajuste automático de los anchos de columna en las tablas pivote:

**Sólo las etiquetas.** Ajusta el ancho de columna al ancho de la etiqueta de la columna. Así se generan tablas más compactas, pero no se mostrarán los valores de los datos más anchos que la etiqueta (los asteriscos indican valores demasiado anchos para ser mostrados).

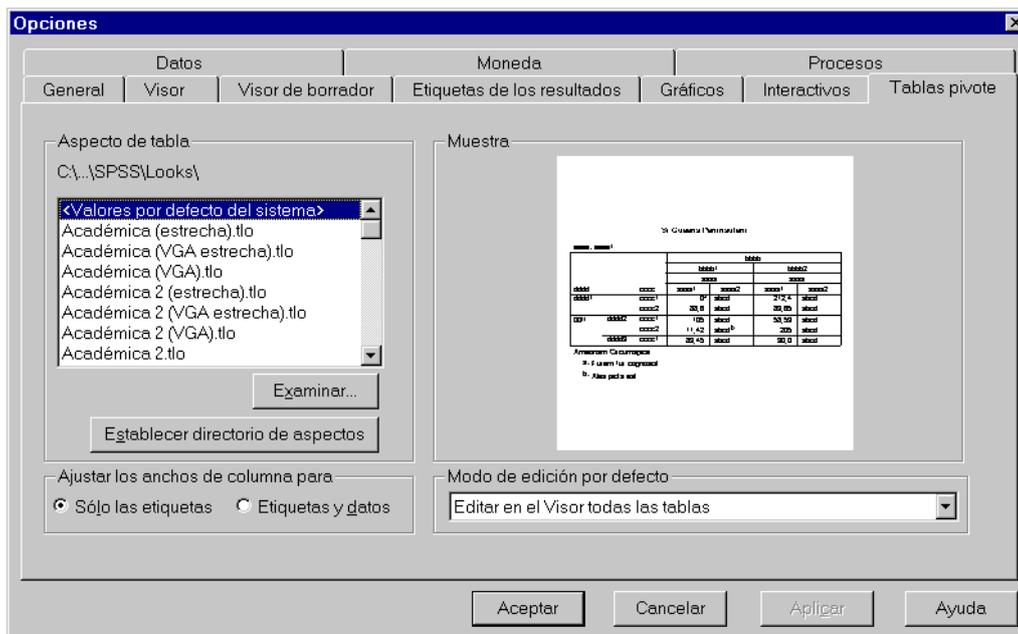


Figura 48: Pestaña de Tablas pivote, dentro de la opción Edición ⇒ Opciones

**Etiquetas y Datos.** Ajusta el ancho de columna al más ancho de entre la etiqueta de columna y el mayor de los valores de los datos. Así se generan tablas más anchas, pero se asegura que se mostrarán todos los valores.

El apartado **Modo de edición por defecto** controla la activación de las tablas pivote en la ventana del visor o en una ventana independiente. Por defecto, cuando se pulsa dos veces en una tabla pivote, ésta se activa en la ventana del visor. Se puede elegir entre activar las tablas pivote en una ventana independiente o seleccionar una opción de tamaño que abra las tablas pivote más pequeñas en la ventana del visor y las más grandes en una ventana independiente.

### 2.2.3. Cómo cambiar las especificaciones del visor de resultados de SPSS por defecto

Para establecer las especificaciones del visor de resultados de SPSS por defecto hay que seleccionar **Edición ⇒ Opciones ⇒ Pestaña de Visor**. Aparece entonces el cuadro de diálogo de la Figura 49. Las opciones de salida del visor sólo afectan a

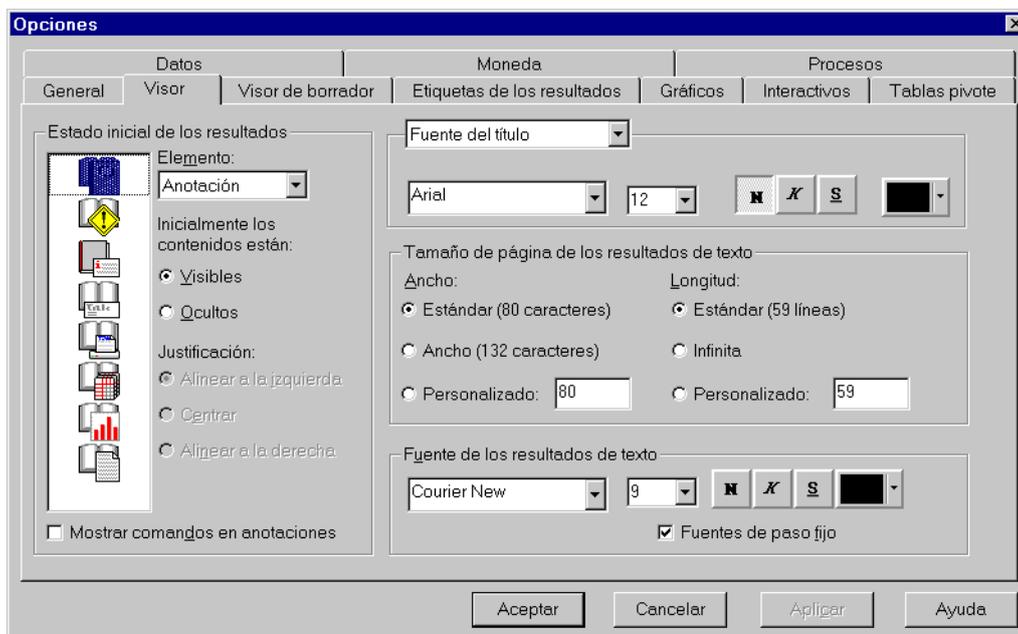


Figura 49: Pestaña de Visor, dentro de la opción Edición ⇒ Opciones

los resultados obtenidos tras el cambio de la configuración. A los resultados que ya se muestran en el visor no les afectan los cambios en la configuración.

El apartado **Estado inicial de los resultados** controla los elementos que se muestran y se ocultan automáticamente cada vez que se ejecuta un procedimiento, además de la alineación inicial de los elementos. Se puede controlar la visualización de los

siguientes elementos: registro, advertencias, notas, títulos, tablas pivote, gráficos y resultados de texto (los resultados que no se muestran en las tablas pivote).

Todos los elementos de resultados aparecen alineados a la izquierda en el visor. Las selecciones de justificación sólo afectarán a la alineación de los resultados impresos. Los elementos con alineación centrada y a la derecha se identifican por un pequeño símbolo situado encima y a la izquierda del elemento.

**Fuente del título** controla el estilo, el tamaño y el color de la fuente de los nuevos títulos de resultados.

**Fuente del título de página** controla el estilo, el tamaño y el color de la fuente de los nuevos títulos de páginas.

**Tamaño de página de los resultados de texto** controla el ancho de página (expresado en número de caracteres) y el largo de página (expresado en número de líneas) en los resultados de texto. En algunos procedimientos, algunos estadísticos se muestran sólo en formato ancho.

**Fuentes de los resultados de texto** controla las fuentes utilizadas para los resultados de texto. Los resultados de texto se han diseñado para utilizarlos con **Fuentes de paso fijo**. Si se selecciona una fuente que no sea de paso fijo, los resultados tabulares no se alinearán adecuadamente.

## 2.2.4. Cómo guardar y recuperar una ventana del visor de resultados de SPSS

Antes de cerrar una ventana del visor de resultados, el sistema nos pregunta si queremos guardar los resultados. La forma de guardar esta ventana de resultados es la habitual en una aplicación Windows; pero la extensión de este documento ha de ser *.spo*. Si guardamos una ventana de resultados podremos recuperarla en cualquier otra sesión. Para ello tenemos que abrir la ventana del visor de resultados (con **Archivo** ⇒ **Nuevo** ⇒ **Resultados**) y en ella seleccionar **Archivo** ⇒ **Abrir**.

## 2.3. Medidas descriptivas de los datos

Como ya hemos dicho, para obtener las medidas descriptivas de una muestra de datos se selecciona la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**. Como también hemos dicho, en primer lugar hay que seleccionar, en el cuadro de diálogo Figura 45, la variable o variables de las cuales se quiere calcular sus medidas descriptivas, y a continuación se pulsa en el botón **Estadísticos**. Entonces aparece el cuadro de diálogo de la Figura 50. En este cuadro tenemos:

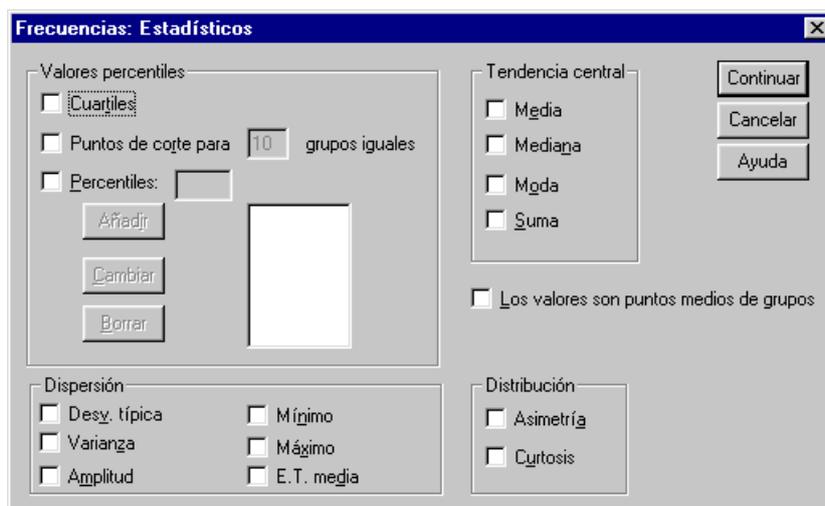


Figura 50: Cuadro de diálogo para seleccionar las medidas descriptivas o estadísticos

**Valores percentiles.** Si se quieren calcular los cuartiles hay que activar **Cuartiles**. Si se desea calcular los valores que dividen la muestra ordenada en  $n$  partes iguales hay que activar **Puntos de corte para**  **grupos iguales**. Por defecto, si se activa esta opción, se calculan los deciles (en  hay un 10). Para obtener otros percentiles hay que activar **Percentiles** y especificar el porcentaje. Para lograr una lista de ellos se pulsa, tras cada uno de ellos, el botón **Añadir**; para cambiar alguno de ellos se pulsa el botón **Cambiar**; y para borrar alguno de los especificados se pulsa el botón **Borrar**.

Hay que tener en cuenta una cuestión importante: el sistema no calcula los percentiles de la misma forma en que lo hemos explicado en clase, pero el cálculo es correcto en ambos casos, pues lo importante es que verifique su definición: un percentil (o cuantil) al  $r$  % es aquel valor que deja a su izquierda el  $r$  % de las observaciones y a su derecha el  $(100 - r)$  % restante.

**Tendencia central.** Los estadísticos o medidas descriptivas de tendencia central que se calculan son: **Media**, **Mediana**, **Moda** y **Suma** de todos los valores.

**Dispersión.** Los estadísticos o medidas descriptivas de dispersión que se calculan son: **Desv. típica** (es, en realidad, la cuasidesviación típica  $S$ ), **Varianza** (es, en realidad, la cuasivarianza  $S^2$ ), **Amplitud** (es el rango o recorrido total  $R = x_{max} - x_{min}$ ), **Mínimo** ( $x_{min}$ ), **Máximo** ( $x_{max}$ ) y **E. T. media** (error típico de la media:  $S/\sqrt{n}$ ).

Una vez hallada la cuasivarianza  $S^2$  (que aquí se llama *Varianza*) podemos hallar la varianza  $s^2$  pues:

$$n s^2 = (n - 1) S^2$$

y por tanto

$$s^2 = \frac{n - 1}{n} S^2 .$$

En consecuencia, una vez hallada la cuasidesviación típica,  $S$ , es posible calcular la desviación típica,  $s$ , mediante:

$$s = \sqrt{\frac{n - 1}{n} S^2} = \sqrt{\frac{n - 1}{n}} S .$$

**Distribución.** Activando las opciones **Asimetría** y **Curtosis** se calculan las medidas de forma: coeficientes de asimetría y curtosis (apuntamiento).

**Los valores son puntos medios de grupos.** Si se han agrupado los datos en intervalos, se pueden estimar los percentiles para los datos originales, es decir, sin agrupar, suponiendo que los casos se distribuyen uniformemente en cada grupo. Por supuesto carece de sentido aplicarlo a variables no agrupadas en intervalos. Se podría aplicar a variables no agrupadas en intervalos considerando que los intervalos tienen una amplitud unitaria (por supuesto, los percentiles no coinciden con los calculados de la forma anterior).

N	Válidos	30
	Perdidos	0
Media		23.820
Error típ. de la media		.923
Mediana		21.500
Moda		20.0
Desv. típ.		5.053
Varianza		25.534
Asimetría		1.259
Error típ. de asimetría		.427
Curtosis		.605
Error típ. de curtosis		.833
Rango		17.4
Mínimo		18.8
Máximo		36.2
Suma		714.6
Percentiles	25	20.000
	50	21.500
	75	25.675

Figura 51: Medidas descriptivas o estadísticos de la variable altura

Si del fichero **Libros.sav** seleccionamos la variable **altura**, en el visor aparecen dos tablas: una con las medidas descriptivas o estadísticos (tabla de la Figura 51) y otra con la distribución de frecuencias (tabla de la Figura 52).

Retomamos ahora los datos del número de palabras por línea de una página de un libro:

palabras por línea	4	5	8	9	10	11	12	13	14	15	16	17
frecuencia absoluta	1	1	2	3	2	7	11	14	3	2	1	1

que estaban grabados en el fichero **PalabrasPorLinea.sav**. Recordemos que en dicho fichero también está grabada la ponderación que hicimos en el apartado 1.15: los datos de la primera columna (que contiene los resultados de la variable) están ponderados mediante la segunda columna (que consta de las correspondientes frecuencias absolutas).

Vamos a obtener las principales medidas descriptivas de la variable **palabras** y la tabla con la distribución de frecuencias. Para ello, en el cuadro de diálogo de la Figura 45 seleccionamos la variable **palabras** para ponerla en el recuadro **Variables**. En este cuadro de diálogo tenemos que dejar activada la opción **Mostrar tablas de frecuencias** pues queremos que aparezca la distribución de frecuencias. En este mismo cuadro de diálogo pulsamos el botón **Estadísticos** y aparece el cuadro de diálogo de la Figura 50. En él seleccionamos los estadísticos o medidas descriptivas que deseamos, por ejemplo, los cuartiles, los deciles,

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	18.8	1	3,3	3,3	3,3
	19.8	2	6,7	6,7	10,0
	20.0	8	26,7	26,7	36,7
	20.1	1	3,3	3,3	40,0
	20.5	2	6,7	6,7	46,7
	21.0	1	3,3	3,3	50,0
	22.0	1	3,3	3,3	53,3
	22.8	1	3,3	3,3	56,7
	23.5	2	6,7	6,7	63,3
	25.0	2	6,7	6,7	70,0
	25.3	1	3,3	3,3	73,3
	25.6	1	3,3	3,3	76,7
	25.9	1	3,3	3,3	80,0
	28.9	1	3,3	3,3	83,3
	30.6	1	3,3	3,3	86,7
	30.9	1	3,3	3,3	90,0
	33.0	1	3,3	3,3	93,3
	35.9	1	3,3	3,3	96,7
	36.2	1	3,3	3,3	100,0
	Total		30	100,0	100,0

Figura 52: Distribución de frecuencias de la variable altura

la media, la cuasivarianza, el valor mínimo y el valor máximo. Por supuesto, la opción **Los valores son puntos medios de grupos** tiene que estar desactivada pues los valores que aparecen en la primera columna del fichero de datos **PalabrasPorLinea.sav** no son las marcas de clase de unos intervalos, sino que son los resultados reales de la variable en la muestra. Tras pulsar el botón **Continuar** en el cuadro de la Figura 50; y el botón **Aceptar** en el de la Figura 45 nos aparece, en la ventana del visor de resultados, la tabla de la distribución de frecuencias (tabla de la Figura 53) y la tabla con los estadísticos seleccionados (tabla de la Figura 54). En este ejemplo, como la cuasivarianza es  $S^2 = 5'76$ , entonces la varianza es:

$$s^2 = \frac{n-1}{n} S^2 = \frac{47}{48} 5'76 = 5'64.$$

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	4	1	2,1	2,1	2,1
	5	1	2,1	2,1	4,2
	8	2	4,2	4,2	8,3
	9	3	6,3	6,3	14,6
	10	2	4,2	4,2	18,8
	11	7	14,6	14,6	33,3
	12	11	22,9	22,9	56,3
	13	14	29,2	29,2	85,4
	14	3	6,3	6,3	91,7
	15	2	4,2	4,2	95,8
	16	1	2,1	2,1	97,9
	17	1	2,1	2,1	100,0
Total		48	100,0	100,0	

Figura 53: Distribución de frecuencias de la variable palabras

**Las medidas descriptivas se pueden obtener también mediante** la opción **Analizar**  $\Rightarrow$  **Estadísticos descriptivos**  $\Rightarrow$  **Descriptivos**. Nos aparece entonces el cuadro de diálogo de la Figura 55.

Si se seleccionan las variables y a continuación se pulsa en **Aceptar**, entonces se calculan sólo las siguientes medidas descriptivas: N (tamaño muestral), Mínimo, Máximo, Media y Desv. típica (cuasidesviación típica). Con el fin de obtener otros estadísticos se pulsa el botón **Opciones...** del cuadro de diálogo de la Figura 55 y entonces aparece el cuadro de la Figura 56.

En este cuadro, lo único que es distinto de lo que aparece en el cuadro de la Figura 50 es lo siguiente:

**Orden de visualización.** Por defecto, las variables se muestran en el orden en que se hayan seleccionado. Si se desea, se pueden mostrar las variables alfabéticamente, por medias ascendentes o por medias descendentes.

N	Válidos	48
	Perdidos	0
Media		11,83
Varianza		5,76
Mínimo		4
Máximo		17
Percentiles	10	8,90
	20	10,80
	25	11,00
	30	11,00
	40	12,00
	50	12,00
	60	13,00
	70	13,00
	80	13,00
	90	14,10

Figura 54: Medidas descriptivas o estadísticos de la variable palabras



Figura 55: Cuadro de diálogo de la opción Analizar ⇒ Estadísticos descriptivos ⇒ Descriptivos

Las diferencias de este procedimiento (**Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Descriptivos**) con el procedimiento anterior (**Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**) son:

- En este procedimiento, en la lista de variables del cuadro de diálogo sólo nos aparecen las que son cuantitativas. En el otro aparecen todas.
- Con este procedimiento, si se eligen varias variables, los resultados se muestran en una misma tabla. Además, las variables se pueden ordenar por el tamaño de sus medias (en orden ascendente o descendente), alfabéticamente o por el orden en el que se seleccionen las variables (el valor por defecto).
- Además, este procedimiento calcula valores tipificados (puntuaciones  $z = [x - \text{Media}] / \text{Desviación típica}$ ). Cuando se guardan las puntuaciones z, éstas se añaden a los datos del *Editor de datos*, quedando disponibles para los gráficos, el listado



Figura 56: Cuadro de diálogo de las Opciones dadas en Analizar ⇒ Estadísticos descriptivos ⇒ Descriptivos

de los datos y los análisis. Cuando las variables se registran en unidades diferentes (por ejemplo, producto interno bruto per cápita y porcentaje de alfabetización), una transformación de puntuación z pondrá las variables en una escala común para una comparación visual más fácil.

## 2.4. Representaciones gráficas unidimensionales

Para obtener las representaciones gráficas de las distribuciones de frecuencias se selecciona la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**. Como hemos dicho, en primer lugar hay que seleccionar, en el cuadro de diálogo Figura 45, la variable o variables de las cuales se quiere obtener su representación gráfica, y a continuación se pulsa en el botón **Gráficos**. Entonces aparece el cuadro de diálogo de la Figura 57. En este cuadro tenemos:

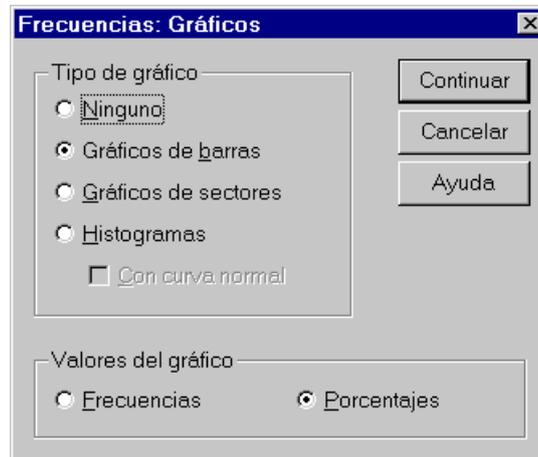


Figura 57: Cuadro de diálogo para seleccionar el tipo de gráfico

**Tipo de gráfico.** Los gráficos de sectores muestran la contribución de las partes a un todo. Cada sector de un gráfico de este tipo corresponde a un grupo, definido por una única variable de agrupación. Los gráficos de barras muestran la frecuencia de cada valor o categoría distinta como una barra diferente, permitiendo comparar las categorías de forma visual. Los histogramas también cuentan con barras, pero se representan a lo largo de una escala de intervalos iguales. La altura de cada barra es el recuento de los valores que están dentro del intervalo para una variable cuantitativa. Una curva normal superpuesta en un histograma ayuda a juzgar si los datos están normalmente distribuidos.

**Valores del gráfico.** Para los gráficos de barras se puede etiquetar el eje vertical con las frecuencias (absolutas) o los porcentajes.

Si del fichero **Libros.sav** seleccionamos la variable **altura**, y del cuadro de diálogo de la Figura 57 elegimos **Gráficos de barras**, obtenemos la representación gráfica de la Figura 58.

Volvamos otra vez a los datos del número de palabras por línea de una página de un libro:

palabras por línea	4	5	8	9	10	11	12	13	14	15	16	17
frecuencia absoluta	1	1	2	3	2	7	11	14	3	2	1	1

que grabados en el fichero **PalabrasPorLinea.sav**.

Vamos a obtener el diagrama de barras de porcentajes de la variable **palabras**. Para ello, en el cuadro de diálogo de la Figura 45 seleccionamos la variable **palabras** para ponerla en el recuadro **Variables**. En este cuadro de diálogo desactivamos la opción **Mostrar tablas de frecuencias** pues no queremos que aparezca la distribución de frecuencias. En este mismo cuadro de diálogo pulsamos el botón **Gráficos** y aparece el cuadro de diálogo de la Figura 57. En el apartado **Tipo de gráfico** seleccionamos **Gráficos de barras**; y en el apartado **Valores del gráfico** elegimos **Porcentajes**. Tras pulsar **Continuar** y **Aceptar** nos aparece, en la ventana del visor de resultados, el gráfico de la Figura 59.

**Observación** El aspecto de cualquier gráfica se puede modificar y mejorar a gusto del usuario (consultar el apartado 2.4.3). Concretamente, las representaciones gráficas de las Figuras 58 y 59 han sido modificadas con el editor de gráficos.

### 2.4.1. Cómo cambiar las especificaciones de las representaciones gráficas por defecto

Ahora que hemos hecho nuestro primer gráfico, quisiera señalar que las especificaciones generales de las representaciones gráficas se pueden cambiar mediante **Edición** ⇒ **Opciones** ⇒ **Pestaña de Gráficos**. Nos aparece entonces el cuadro de diálogo de la Figura 60. Los distintos apartados de este cuadro de diálogo son:

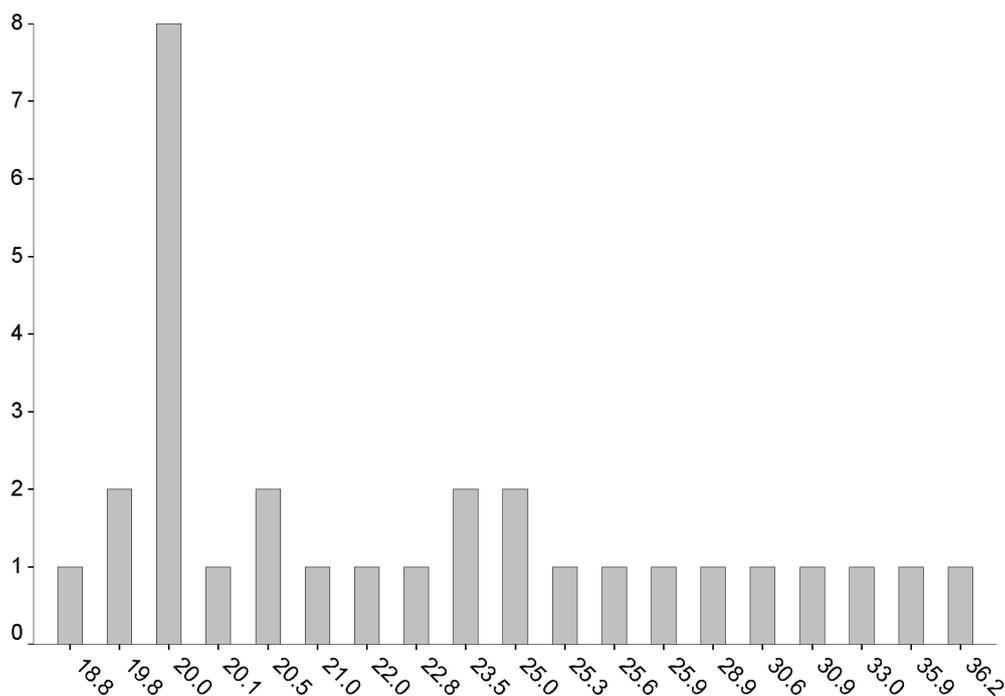


Figura 58: Diagrama de barras de frecuencias absolutas de la variable altura

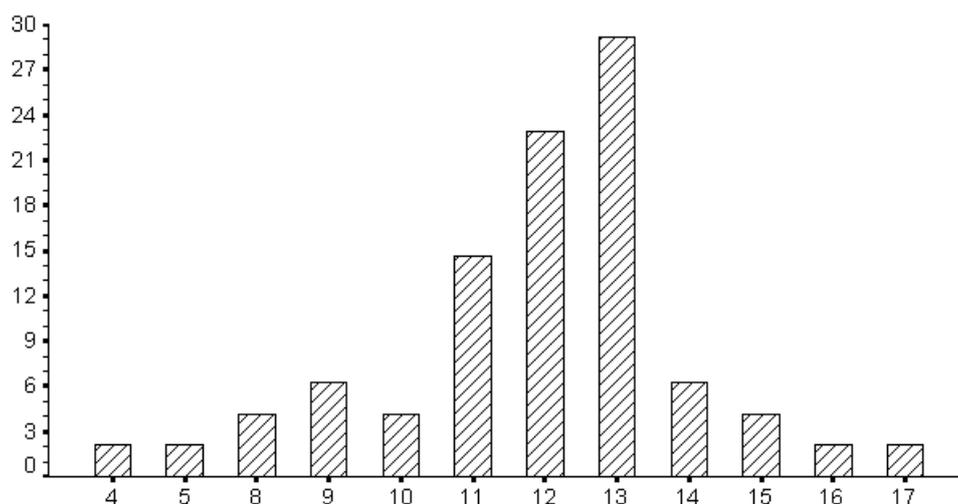


Figura 59: Diagrama de barras de porcentajes de la variable palabras

**Plantilla gráfica.** Los gráficos nuevos pueden utilizar tanto las opciones seleccionadas aquí, como las opciones de un archivo de plantilla gráfica. Se puede pulsar en **Examinar** para seleccionar un archivo de plantilla gráfica.

**Relación de aspecto del gráfico.** Este apartado sirve para fijar la relación ancho-alto del marco exterior de los nuevos gráficos. Se puede especificar una relación ancho-alto entre los valores 0'1 y 10. Los valores inferiores a 1 generan gráficos que son más altos que anchos. Los valores mayores que 1 producen gráficos que son más anchos que altos. Un valor de 1 produce un gráfico cuadrado. (La relación que yo habitualmente utilizo es de 1'5 sino no hay leyendas en los gráficos, y de 1'8 si las hay.)

**Fuente.** Aquí se fija la fuente utilizada para todo el texto en los nuevos gráficos. Por defecto es *Arial*.

**Tramas de relleno y estilos de línea.** Se hace la asignación inicial de colores y tramas para nuevos gráficos:

**Recorrer los colores y después las tramas** utiliza la paleta por defecto de 14 colores y luego añade tramas a los colores si es necesario.

**Recorrer las tramas** sólo utiliza tramas para diferenciar los elementos de los gráficos y no utiliza color.

**Marcos.** Controla la visualización de los marcos interno y externo en los nuevos gráficos.

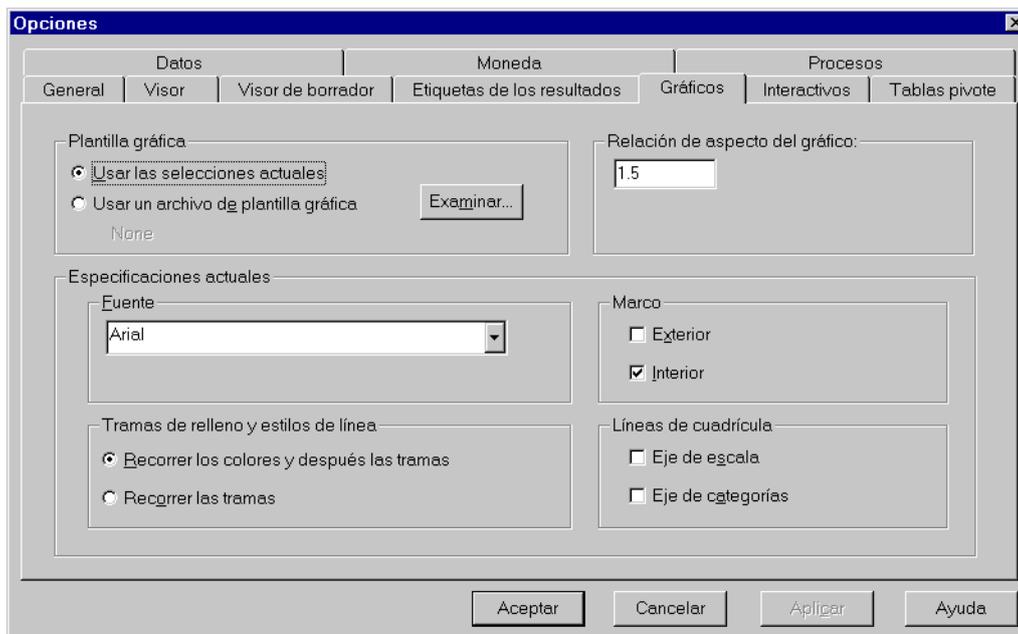


Figura 60: Pestaña de Gráficos, dentro de la opción Edición ⇒ Opciones

**Líneas de cuadrícula.** Controla la visualización de las líneas de cuadrícula en el eje de escala (eje vertical) y en el eje de categorías (eje horizontal) de los nuevos gráficos.

#### 2.4.2. Creación de diagramas de barras, líneas, áreas y sectores

Además de con la opción **Analizar ⇒ Estadísticos descriptivos ⇒ Frecuencias**, las representaciones gráficas también se pueden obtener a partir del menú **Gráficos** de la barra de menús de **SPSS**. Este menú brinda muchas posibilidades, por lo que voy a dedicar esta sección y las tres siguientes a explicar esta opción con más detenimiento.

En esta sección se van a dar con más detalle las explicaciones correspondientes a los diagramas de barras pues las posibilidades de los otros tres tipos de diagramas (líneas, áreas y sectores) son similares.

Para obtener el diagrama de barras de la distribución de frecuencias de una variable se selecciona **Gráficos ⇒ Barras**. Con esta opción aparece el cuadro de diálogo de la Figura 61.



Figura 61: Cuadro de diálogo para hacer un diagrama de barras mediante Gráficos ⇒ Barras

Las alternativas que se ofrecen se refieren a dos aspectos:

- En la mitad superior de cuadro de diálogo se nos presentan tres tipos de iconos según la forma general del diagrama de barras:

**Simple.** Las barras representan elementos del mismo conjunto.

**Agrupado.** Las barras, agrupadas, representan elementos de distintos conjuntos.

**Apilado.** Las barras, apiladas, representan elementos de distintos conjuntos.

- En la mitad inferior de cuadro de diálogo, en **Los datos del gráfico son**, se puede seleccionar la estructura de los datos:

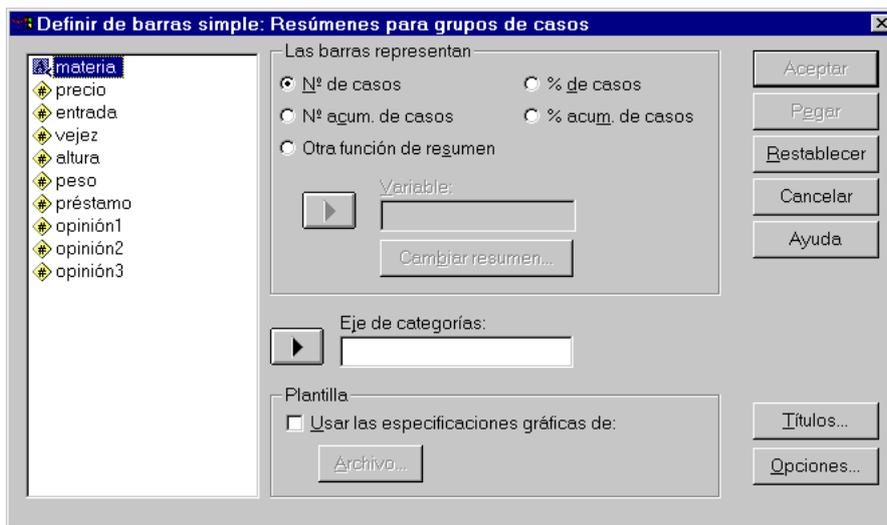


Figura 62: Cuadro de diálogo para hacer un diagrama de barras Simple con la opción Resúmenes para grupos de casos

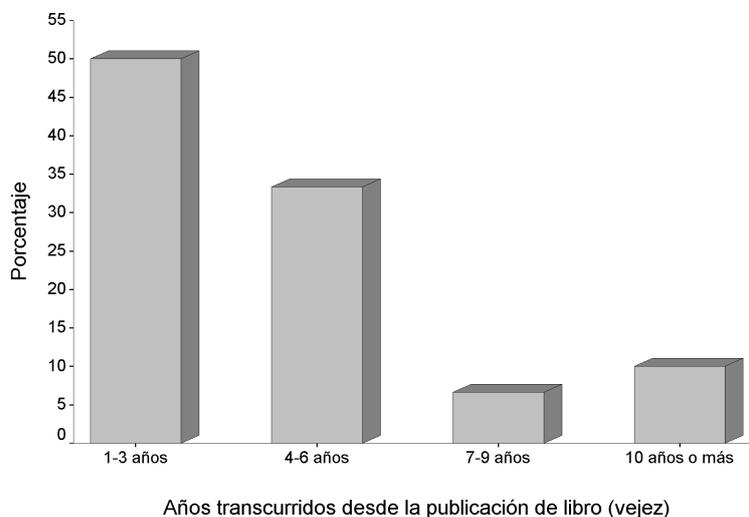


Figura 63: Diagrama de barras de porcentajes de la variable vejez

**Resúmenes para grupos de casos.** Cada barra representa la frecuencia de cada categoría de una variable.

Si, por ejemplo, elegimos el icono **Simple** y en *Los datos del gráfico son* activamos la opción **Resúmenes para grupos de casos**, entonces aparece el cuadro de diálogo de la Figura 62. En este cuadro de diálogo, en **Eje de categorías** se selecciona la variable cuya distribución de frecuencias se desea representar. Por defecto, haciendo *click* en **Aceptar**, se obtiene la gráfica en que cada barra representa la frecuencia absoluta de cada categoría. Sin embargo, se pueden seleccionar otros tipos de frecuencias en **Las barras representan**:

**Nº de casos.** Frecuencia absoluta de cada categoría.

**Nº acumulado de casos.** Frecuencia absoluta acumulada de cada categoría.

**% de casos.** Porcentaje de casos de cada categoría sobre el total.

**% acumulado de casos.** Porcentaje acumulado de casos de cada categoría.

**Otra función de resumen.** El eje vertical, así como la altura de cada barra se asocia a otra variable que es especificada en el recuadro **Variable**. Haciendo *click* en el botón **Cambiar resumen** se accede al cuadro de diálogo de la función de resumen, que se examinará en el apartado 2.4.6.

Por ejemplo, en el cuadro de diálogo de la Figura 61 seleccionamos el icono **Simple**; en *Los datos del gráfico son* activamos la opción **Resúmenes para grupos de casos**; en **Eje de categorías** ponemos la variable **vejez** (años transcurridos desde la publicación del libro); y en **Las barras representan** seleccionamos **% de casos**, entonces obtenemos el diagrama de barras de porcentajes correspondiente a la variable *vejez*, que es el gráfico de la Figura 63.

Como novedad sobre lo ya expuesto, los diagramas de barras agrupadas y apiladas ofrecen la posibilidad de seleccionar una nueva variable para determinar las barras dentro de cada grupo o de cada apilamiento (ver la Figura 64); esto se realiza a través de **Definir grupos por** (o **Definir las pilas por**) respectivamente, apareciendo tal variable y sus valores en la leyenda adjunta al diagrama ya obtenido.

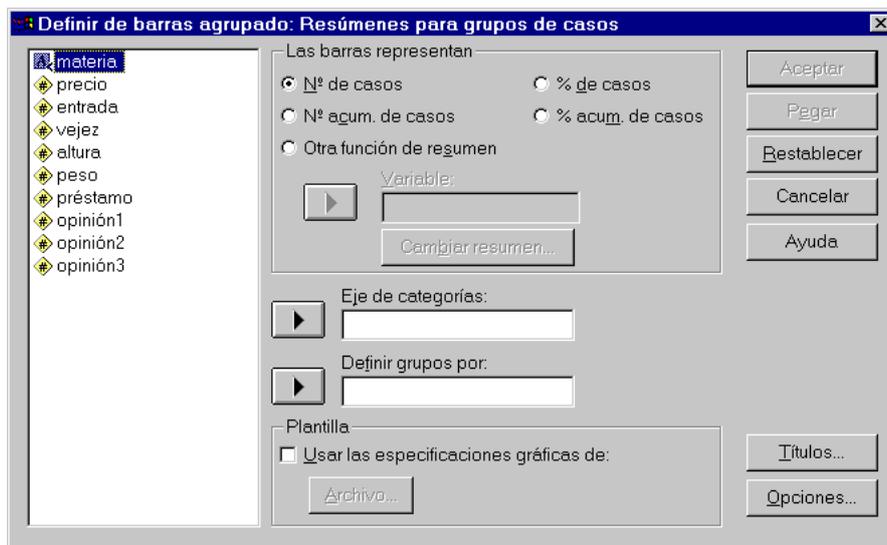


Figura 64: Cuadro de diálogo para hacer un diagrama de barras Agrupado con la opción Resúmenes para grupos de casos

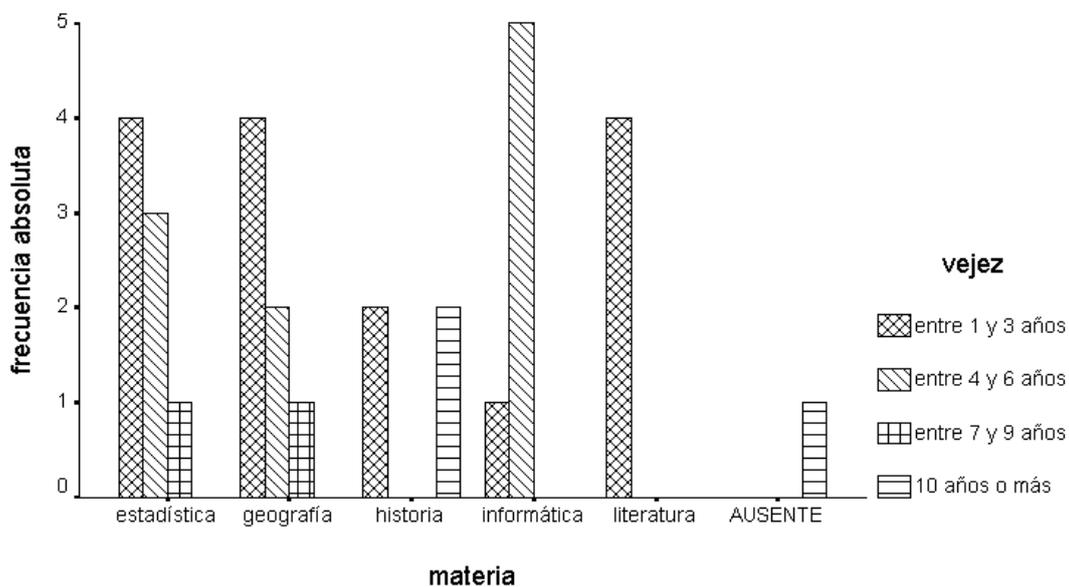


Figura 65: Diagrama de barras de la variable materia en grupos definidos por la variable vejez

Por ejemplo, en el cuadro de diálogo de la Figura 61 seleccionamos el icono **Agrupado**; en *Los datos del gráfico* son activamos la opción **Resúmenes para grupos de casos**; en **Eje de categorías** ponemos la variable **materia** (disciplina científica del libro); y en **Definir grupos por** seleccionamos la variable **vejez** (número de años transcurridos desde la publicación del libro), entonces obtenemos el diagrama de barras de frecuencias absolutas de la Figura 65.

Si ahora hacemos lo mismo cambiando la opción **Agrupado** por **Apilado** entonces obtenemos el diagrama de barras de frecuencias absolutas de la variable **materia** con barras apiladas en grupos definidos por la variable **vejez**, que es el gráfico de la Figura 66.

**Resúmenes para distintas variables.** Cada barra representa una misma modalidad de dos o más variables. Para poder seleccionar esta opción es necesario que los resultados de esas variables estén clasificados en las mismas categorías.

Si elegimos el icono **Simple** y en *Los datos del gráfico* son activamos la opción **Resúmenes para distintas variables**, entonces debemos seleccionar las variables para colocarlas en el recuadro **Las barras representan** (ver Figura 67). Esto se consigue haciendo doble *click* sobre el nombre de dichas variables. Por defecto, haciendo *click* en **Aceptar**, cada barra representa la media aritmética de cada variable. Si queremos cambiar la función que representa cada barra, debemos seleccionar (una a una) las variables que están en el recuadro de *Las barras representan* y después pulsar el botón **Cambiar resumen** (ver el apartado 2.4.6). No es preciso seleccionar el mismo estadístico para todas las variables, aunque esta posibilidad no tenga excesivo sentido.

Por ejemplo, un diagrama de este tipo con las variables **opinión1**, **opinión2** y **opinión3** es el de la Figura 68. En este gráfico se han modificado las funciones de resumen para que aparezca la suma de todos los resultados de

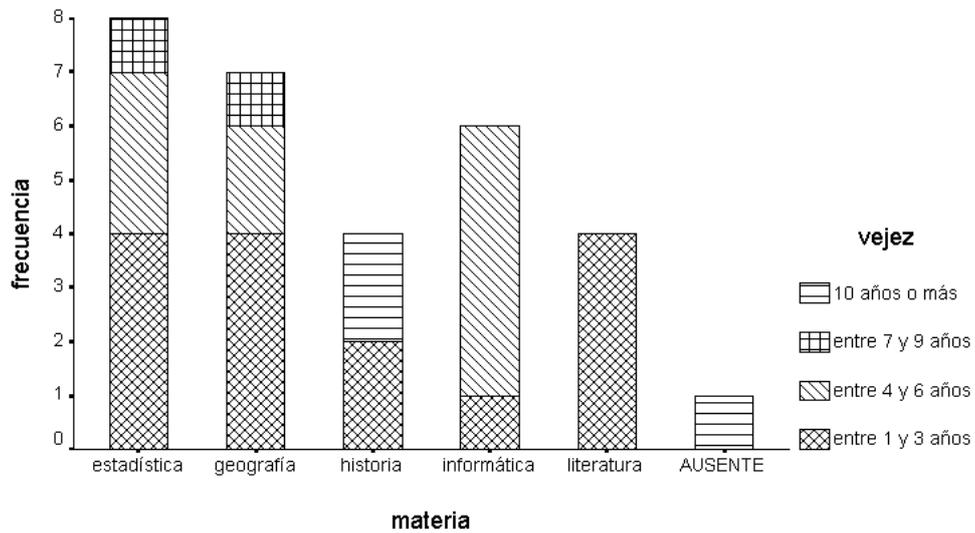


Figura 66: Diagrama de barras de la variable materia apiladas en grupos definidos por la variable vejez

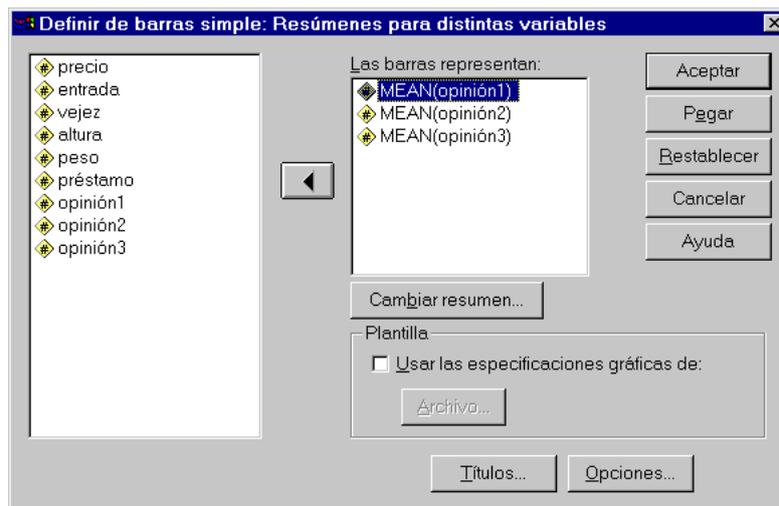


Figura 67: Cuadro de diálogo para hacer un diagrama de barras Simple con la opción Resúmenes para distintas variables

cada variable en vez de la media aritmética (pulsar **Cambiar resumen** y elegir **Suma de valores**).

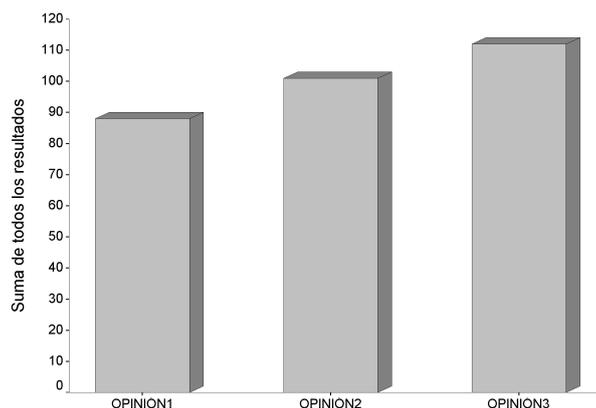


Figura 68: Diagrama de barras con la suma de los resultados de las tres variables de opinión

En los diagramas de barras agrupadas y apiladas se puede (igual que antes) seleccionar, en **Eje de categorías**, una nueva variable a partir de la cual se crea un grupo o un apilamiento de barras. En este caso, en la leyenda adjunta al diagrama aparecen las variables seleccionadas en **Las barras representan**.

**Valores individuales de los casos.** Cada barra representa un caso, lo que quiere decir que en muestras grandes la gráfica sería muy difícil de interpretar.

Si en el cuadro de diálogo de la Figura 61 elegimos **Simple** y **Valores individuales de los casos** entonces aparece el cuadro de diálogo de la Figura 69. En este cuadro, en **Las barras representan** se selecciona una variable en el caso de un diagrama simple y el resultado es que cada barra representa el valor que cada caso toma en esa variable. Esto quiere decir que tal variable debe ser cuantitativa.

En los casos de diagramas agrupados o apilados hay que seleccionar, evidentemente, más de una variable.

En **Etiquetas de las categorías** se puede especificar la etiqueta de cada barra:

**Número del caso.** Por defecto, cada barra es etiquetada por el número del caso al que representa.

**Variable.** Cada barra es etiquetada por el valor que su correspondiente caso adopta en la variable ahora declarada.

En la leyenda adjunta al diagrama se muestran los nombres de las variables en los casos de diagramas agrupados o apilados.

Por ejemplo, si seleccionamos la variable **préstamo** en **Las barras representan** y en **Etiquetas de las categorías** dejamos la opción por defecto (**Número del caso**) obtenemos el diagrama de barras de la Figura 70.

Se puede observar que en el cuadro de diálogo inicial (Figuras 62, 64 y 67), excepto cuando se trata de los valores individuales de los casos (Figura 69), se encuentran dos botones, uno referido a **Títulos** y el otro a **Opciones**; así como un bloque titulado **Plantilla**, que para prácticamente todos los tipos de gráficas son análogos. Estas tres cuestiones se tratarán en los apartados 2.4.7, 2.4.8 y 2.4.9, respectivamente.

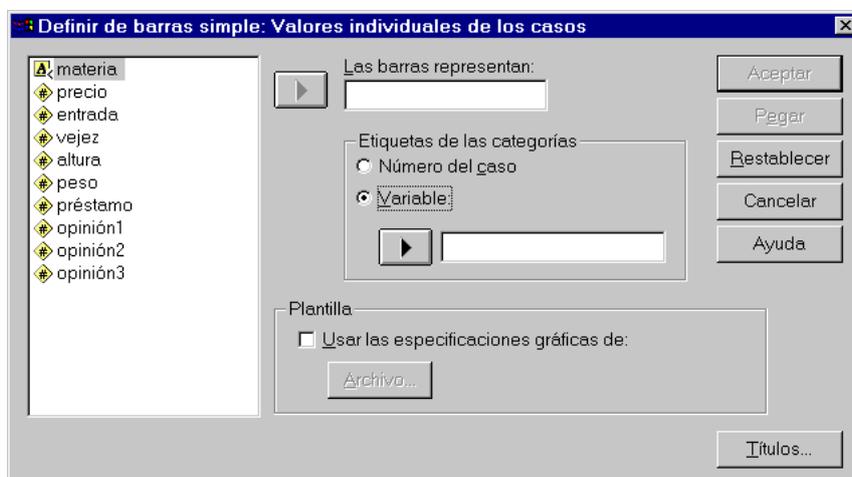


Figura 69: Cuadro de diálogo para hacer un diagrama de barras Simple con la opción Valores individuales de los casos

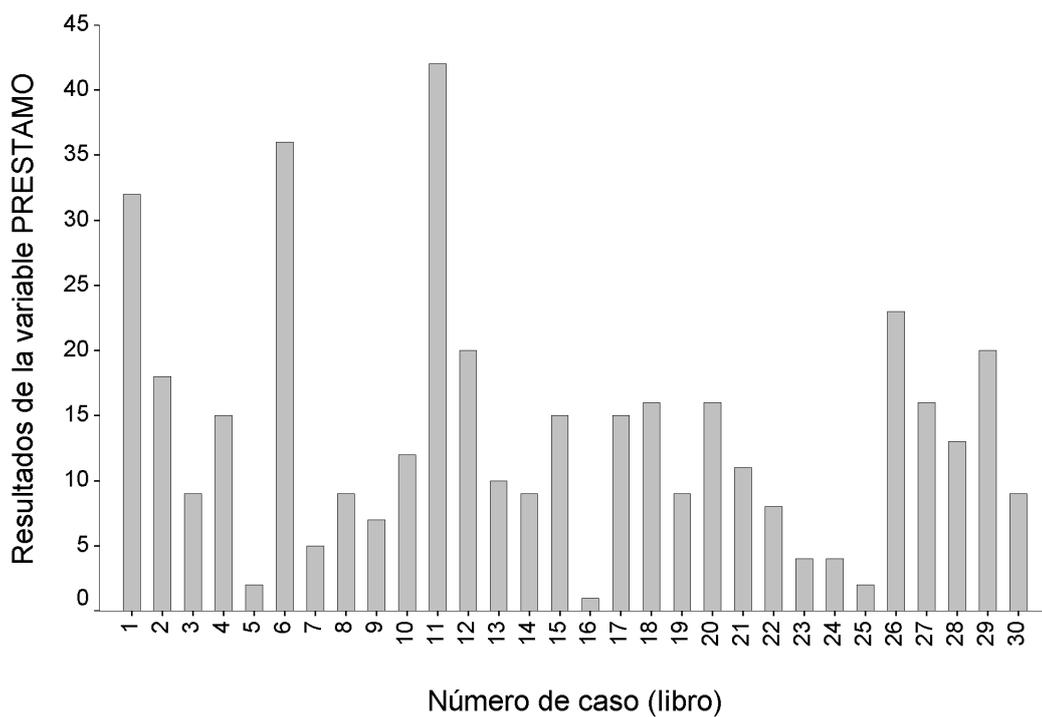


Figura 70: Las barras representan los resultados de la variable préstamo para cada caso (libro)

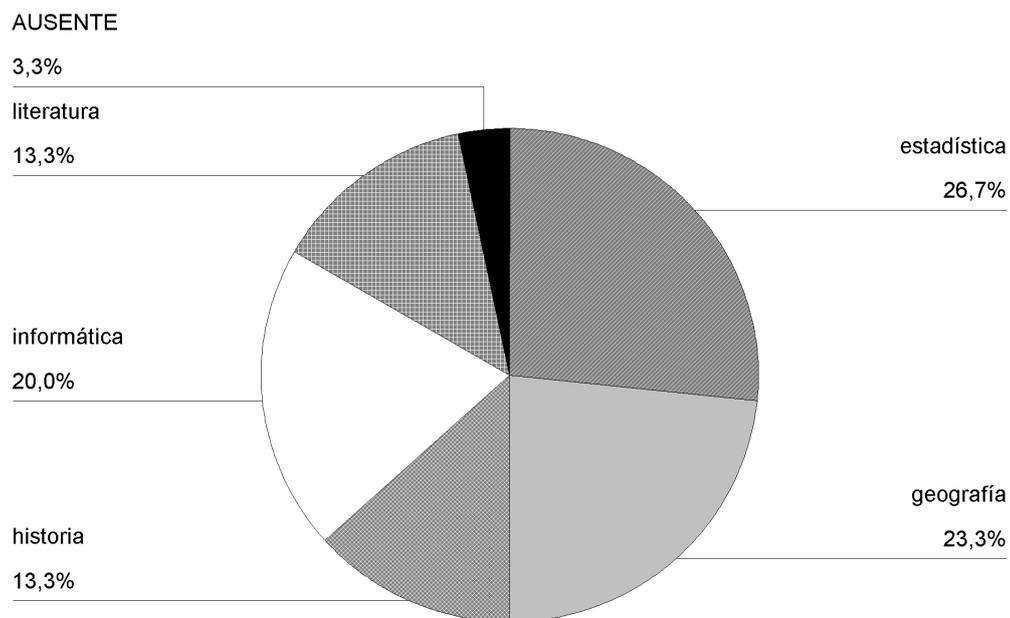


Figura 71: Diagrama de sectores de la variable materia

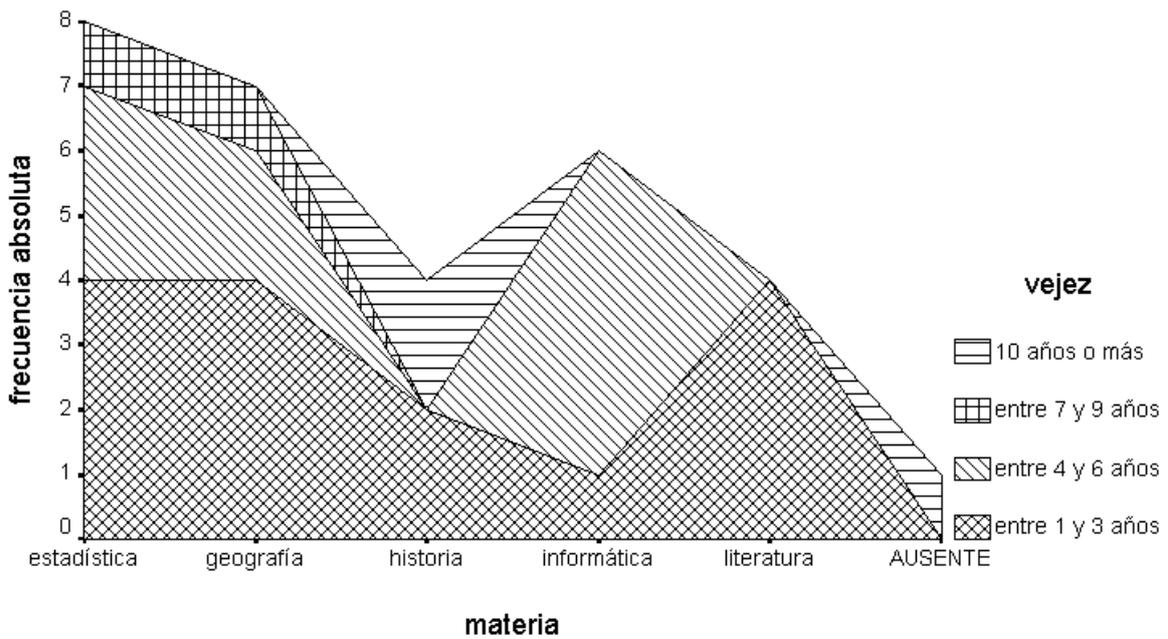


Figura 72: Diagrama de áreas de la variable materia apiladas según los grupos definidos por la variable vejez

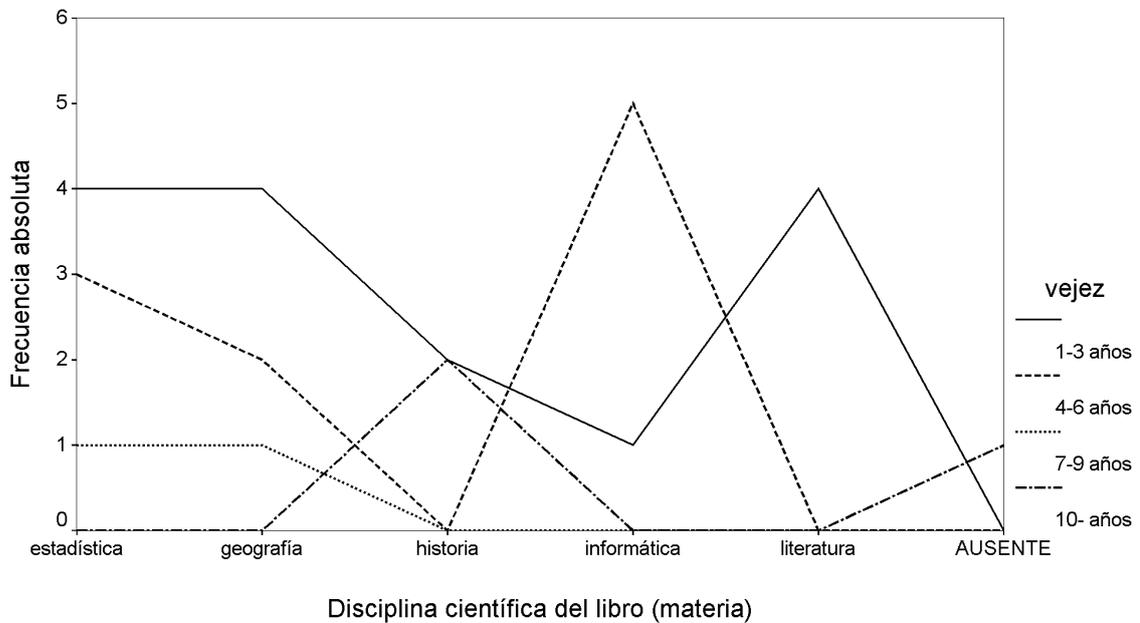


Figura 73: Diagrama múltiple de líneas de la variable materia según los grupos definidos por la variable vejez

Los diagramas de líneas, áreas y sectores se realizan de forma totalmente análoga a los de barras. De todas formas, se puede construir siempre el diagrama de barras y luego se puede utilizar el editor de gráficos (ver el apartado siguiente) para cambiar a cualquier otro tipo de gráfico (con el menú **Galería** del editor de gráficos). Por ejemplo, el diagrama de sectores de la variable **materia** es el de la Figura 71. El diagrama de áreas de la variable **materia** apiladas según los grupos definidos por la variable **vejez** es el gráfico de la Figura 72.

El diagrama múltiple de líneas de la variable **materia** según los grupos definidos por la variable **vejez** es el gráfico de la Figura 73.

### 2.4.3. Edición y modificación de diagramas de barras, líneas, áreas y sectores

Igual que en el apartado anterior, se van a dar con más detalle las explicaciones correspondientes a los diagramas de barras. La extrapolación a los otros tipos de diagramas es sencilla.

Una vez obtenido el diagrama de barras y para proceder a su modificación hay que editarlo haciendo doble *clic* dentro del gráfico en el visor de resultados de **SPSS**. Al hacerlo, entramos en la *ventana del editor de gráficos*. Una vez que estamos en el editor de gráficos se pueden realizar, entre otras, las siguientes modificaciones.



Figura 74: Cuadro para cambiar el estilo de las barras de un diagrama de este tipo



Figura 75: Cuadro para cambiar la trama de relleno de un diagrama de barras

Para modificar el estilo de las barras hay que seleccionar **Formato** ⇒ **Estilos de barra** y en el cuadro de diálogo que aparece (Figura 74) elegir el deseado (normal, sombreado o con efecto de tres dimensiones).

Para cambiar la trama de relleno se selecciona la opción **Formato** ⇒ **Trama de relleno**. Entonces aparece el cuadro de la Figura 75. En él se hace *click* sobre la trama deseada y después se pulsa **Aplicar**. Para terminar este proceso hay que pulsar en **Cerrar**.

Si queremos cambiar el color de cualquier elemento de un diagrama de barras, seleccionamos dicho elemento haciendo *click* sobre él, y después elegimos la opción **Formato** ⇒ **Color**. En el cuadro de diálogo que resulta (Figura 76) se activa **Relleno** o **Borde**, y se selecciona el color deseado haciendo *click* sobre dicho color. Cada vez que modifiquemos el color de un elemento, tenemos que pulsar en **Aplicar**. Para hacer efectivas todas las modificaciones del color hay que hacer *click* en **Cerrar**.

Para modificar el texto y las etiquetas del eje horizontal tenemos que hacer doble *click* sobre dicho eje (o seleccionar la opción **Diseño** ⇒ **Ejes** ⇒ **Categoría** de la barra de menús del editor de gráficos) y luego hacer los cambios deseados en el cuadro de diálogo que aparece (Figura 77). En este cuadro de diálogo hay que pulsar el botón **Etiquetas** para poder cambiar las etiquetas de las categorías o de los resultados de la variable representada en el eje horizontal. Al pulsar este botón aparece el cuadro de la Figura 78.

Análogamente, para modificar el texto y las etiquetas del eje vertical tenemos que hacer doble *click* sobre dicho eje (o seleccionar la opción **Diseño** ⇒ **Ejes** ⇒ **Escala** de la barra de menús del editor de gráficos) y luego hacer los cambios deseados en el cuadro de diálogo que aparece (Figura 79). En este cuadro de diálogo hay que pulsar el botón **Etiquetas** para poder cambiar las etiquetas del eje vertical. Al pulsar este botón aparece el cuadro de la Figura 80.

La opción **Estilos de etiquetas de barras** del menú **Formato** permite que aparezca una etiqueta dentro de cada barra. Esta etiqueta muestra el resultado de lo medido en el eje vertical para cada barra (ver Figura 81). Si se utilizan tramas o colores oscuros, el estilo de etiqueta *enmarcado* proporcionará un área clara en la que la etiqueta podrá leerse con mayor facilidad. Las etiquetas de las barras se muestran con el mismo número de decimales que las etiquetas del eje de escala (eje vertical). Puede cambiarse el número de decimales haciendo *click* dos veces en el eje vertical y, a continuación, eligiendo **Etiquetas** en el cuadro de diálogo **Eje de escala** (ver Figura 80).

Si queremos modificar la fuente y el tamaño de cualquiera de los textos que aparecen en la gráfica, debemos seleccionar dicho texto (haciendo *click* sobre él) y después elegir la opción **Formato** ⇒ **Texto** (ver Figura 82).



Figura 76: Cuadro para cambiar el color de un elemento de un diagrama de barras

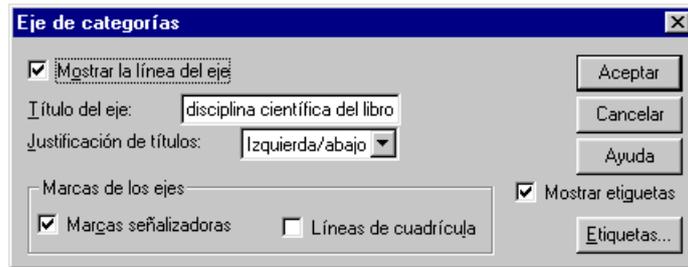


Figura 77: Cuadro para modificar el texto del eje horizontal de un diagrama de barras

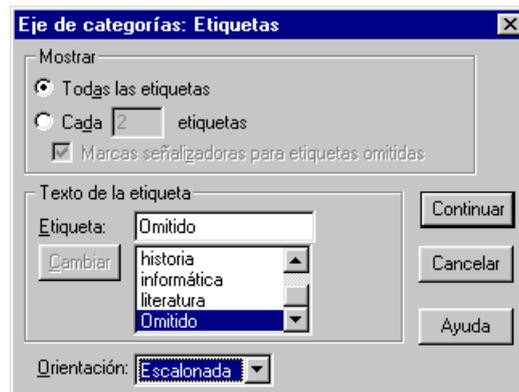


Figura 78: Cuadro para editar las etiquetas del eje horizontal de un diagrama de barras

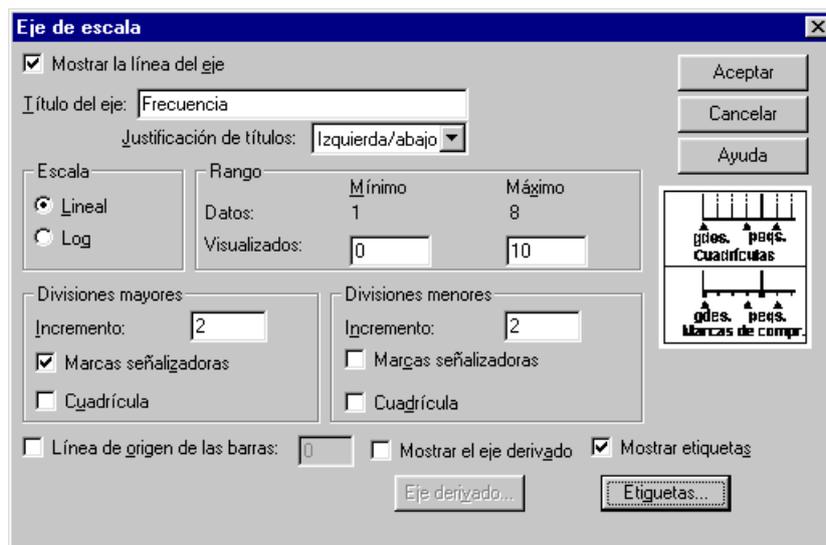


Figura 79: Cuadro para modificar el texto y las etiquetas del eje vertical de un diagrama de barras

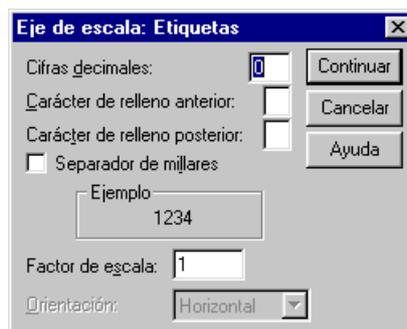


Figura 80: Cuadro para cambiar las etiquetas del eje vertical de un diagrama de barras

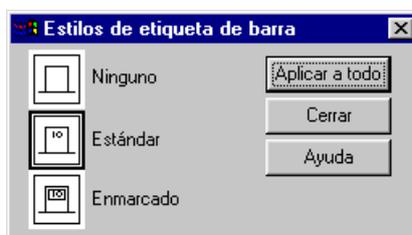


Figura 81: Cuadro para que aparezcan etiquetas en las barras de un diagrama de este tipo



Figura 82: Cuadro para cambiar un texto de un diagrama de barras

Para modificar el espaciado entre las barras debe seleccionarse **Diseño** ⇒ **Espaciado de barras** y cambiar las especificaciones del cuadro de diálogo que aparece (Figura 83).

Si el gráfico lleva una leyenda (por ejemplo, en el caso de un diagrama de barras apiladas o agrupadas) y queremos modificarla, entonces tenemos que hacer doble *clic* sobre ella para que aparezca el cuadro de diálogo de la Figura 84. También podemos seleccionar la opción **Diseño** ⇒ **Leyenda** de la barra de menús del editor de gráficos. En este cuadro de diálogo podemos modificar el título, la justificación (a la izquierda, a la derecha o centrada) y las etiquetas de la leyenda.

Análogamente, las opciones **Título**, **Nota a pie de página** y **Anotación** del menú **Diseño** permiten crear y modificar un título de la gráfica, una nota a pie de página y una anotación de la misma, respectivamente.

La opción **Líneas de referencia** del menú **Diseño** permite añadir líneas de referencia a los gráficos existentes, tanto en el eje horizontal (eje de categorías) como en el eje vertical (eje de escala). Si se desea ocultar la línea de referencia de forma que no esté visible, se selecciona **Ocultar**. Es posible añadir, cambiar y eliminar líneas de referencia.

Además, en el menú **Diseño** se pueden activar y desactivar las opciones **Marco exterior** y **Marco interior**. El *Marco interior* contiene sólo las barras. El *Marco exterior* contiene todos los elementos de la gráfica (barras, ejes, etiquetas, título, nota a pie de página, leyenda, anotación, etc.).

Para intercambiar los ejes se selecciona **Formato** ⇒ **Intercambiar ejes**. Así, entre otras cosas, ocurrirá que las barras verticales se cambiarán por barras horizontales.

#### 2.4.4. Creación de histogramas

Los histogramas no tienen las posibilidades gráficas que ofrecen los diagramas de barras. Además, no es posible obtener el histograma de frecuencias acumuladas ni los polígonos de frecuencias con datos agrupados en intervalos.

Para realizar un histograma se selecciona la opción **Gráficos** ⇒ **Histograma**. Entonces aparece el cuadro de diálogo de la Figura 85. Tras seleccionar una sola variable (cuantitativa) en el recuadro **Variable**, con sólo hacer *clic* en **Aceptar** se obtienen los resultados por defecto.

A la gráfica se puede superponer la curva normal con la misma media y desviación típica que nuestra distribución de frecuencias. Para ello se activa la opción **Mostrar curva normal**.

Por ejemplo, el histograma correspondiente a la variable **precio** es el de la Figura 86. (Este gráfico ha sido ligeramente modificado en la ventana del editor de gráficos -ver el apartado siguiente-).

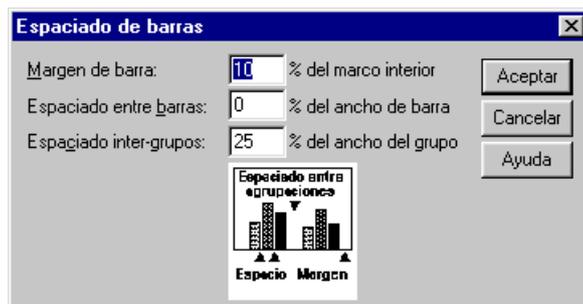


Figura 83: Cuadro para modificar el espaciado entre las barras de un diagrama de este tipo

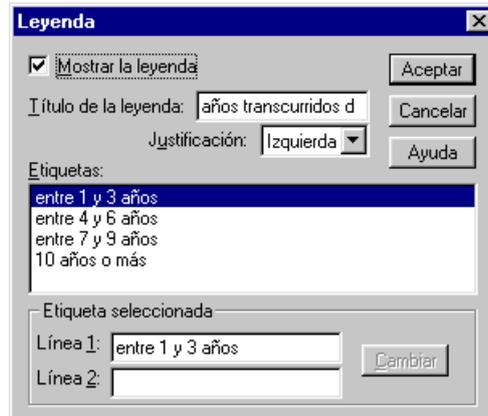


Figura 84: Cuadro para cambiar la leyenda de un diagrama de barras

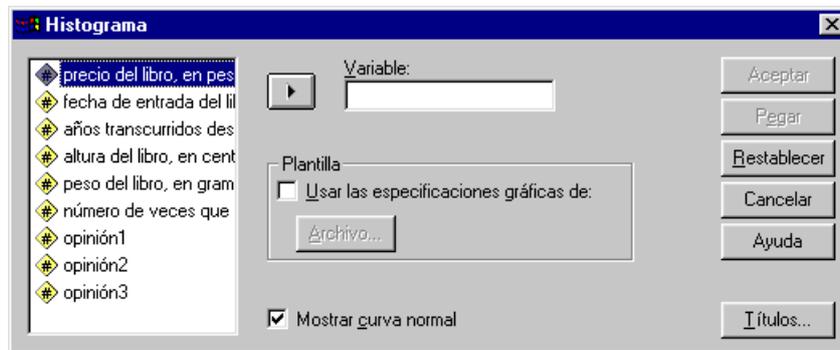


Figura 85: Cuadro de diálogo para hacer un histograma

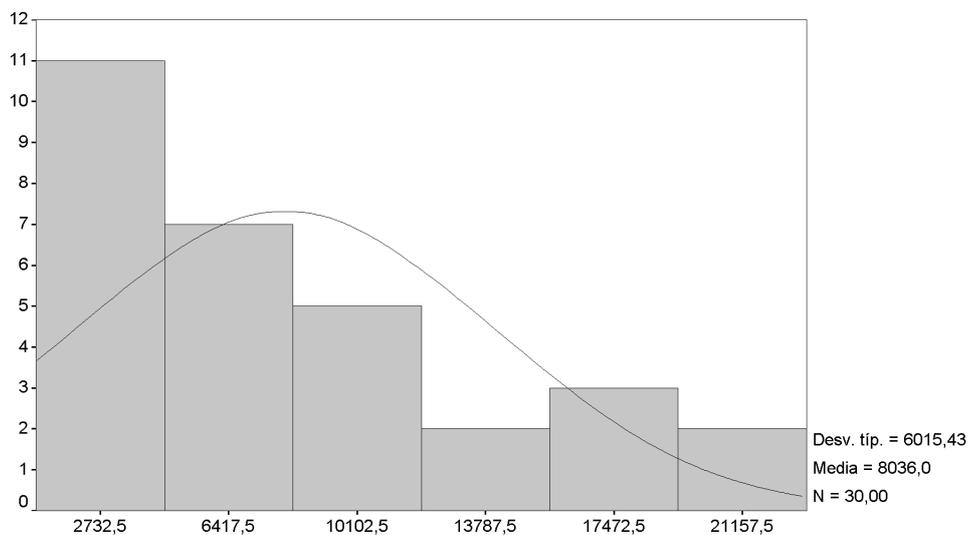


Figura 86: Histograma correspondiente a la variable precio

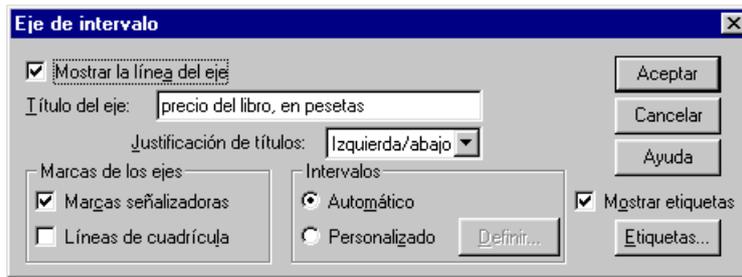


Figura 87: Cuadro para cambiar los intervalos de un histograma



Figura 88: Cuadro para definir los nuevos intervalos de un histograma

#### 2.4.5. Edición y modificación de histogramas

Una vez obtenido el histograma y para proceder a su modificación hay que editarlo (haciendo doble *click* sobre el histograma en el visor de resultados). Las posibilidades de modificación del histograma son muy similares a las del diagrama de barras, por lo que voy a exponer sólo lo que de particular tiene este tipo de gráfico.

En primer lugar, hay que decir que mientras que en el diagrama de barras al eje horizontal se le llamaba eje de categorías, ahora se le llama eje de intervalo.

Es importante también señalar que es posible cambiar los intervalos en se divide el rango de valores de la variable. Para ello se hace doble *click* sobre dicho eje (o se selecciona **Diseño** ⇒ **Ejes** ⇒ **Intervalo** ⇒ **Aceptar**) y en el cuadro de diálogo que aparece (Figura 87) se activa la opción **Personalizado** del grupo **Intervalos** y después se pulsa en **Definir**. Aparece entonces el nuevo cuadro de diálogo de la Figura 88. En este cuadro se tiene que seleccionar una de las opciones del grupo **Definición** para indicar de qué manera se desea especificar los intervalos. Es posible especificar tanto el número de intervalos como la amplitud de los mismos. El grupo **Rango** permite especificar el rango de los valores que se muestran en el eje. Los valores no incluidos en el rango no se muestran en los histogramas.

Pulsando en el botón **Etiquetas** del cuadro de diálogo de la Figura 87 aparece el cuadro de la Figura 89. En él se selecciona una de las opciones del grupo **Mostrar** para controlar la visualización de las etiquetas de los ejes. La opción **Todas las etiquetas** muestra una etiqueta por cada uno de los intervalos representados en el eje horizontal. La opción **Cada n-ésima etiqueta** permite especificar un incremento que determine qué intervalos tendrán etiqueta. Si se ha optado por omitir las etiquetas, es posible desactivar la selección de **Marcas de comprobación para etiquetas omitidas** para que los intervalos sin etiqueta no muestren marcas de comprobación. El grupo **Tipo** permite controlar la manera en que se etiquetan los intervalos. La opción **Punto medio** etiqueta todos los intervalos con sus respectivos puntos medios. La opción **Rango** etiqueta los intervalos con el rango de valores que

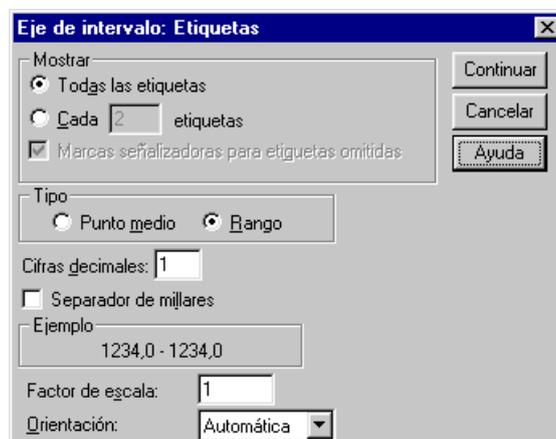


Figura 89: Cuadro para definir las etiquetas del eje horizontal de un histograma

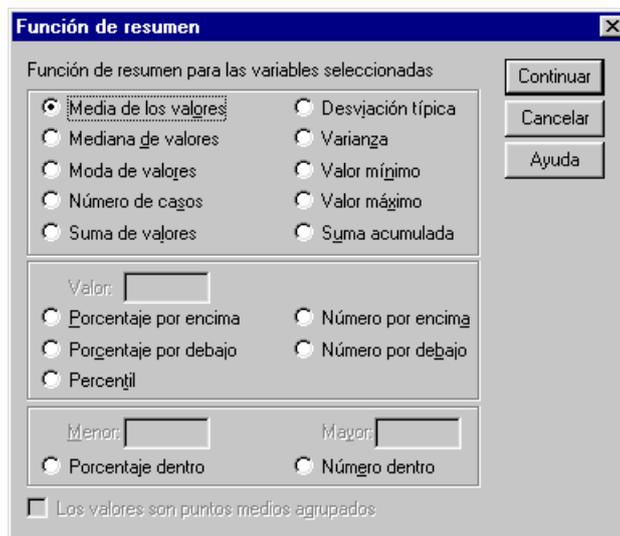


Figura 90: Cuadro de diálogo para cambiar la Función resumen

abarca cada intervalo. Para controlar la cantidad de dígitos decimales de las etiquetas de los ejes, se introduce el valor deseado en **Cifras decimales**. Para utilizar comas o puntos como separador de unidades de millar se selecciona **Separador de millares**. Es posible cambiar el factor de escala introduciendo un nuevo valor en **Factor de escala**. La lista **Orientación** permite controlar la orientación de las etiquetas (horizontal, vertical, diagonal, escalonada).

#### 2.4.6. Otras funciones sintéticas

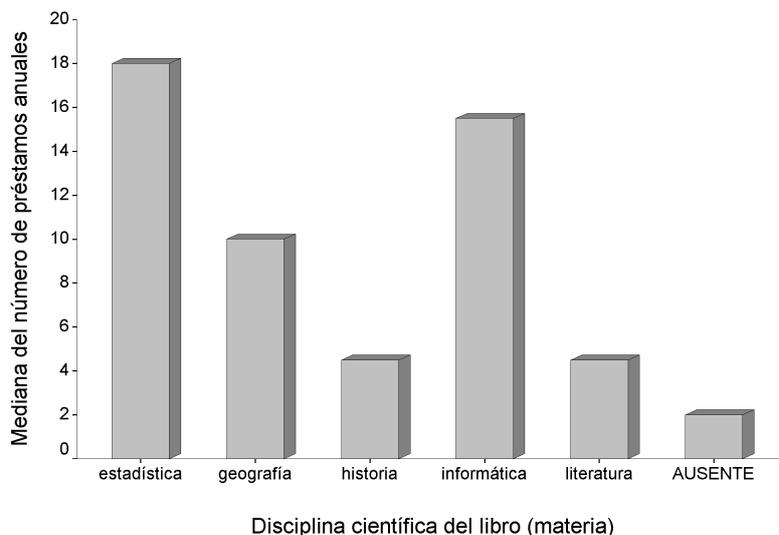


Figura 91: Diagrama de barras de la variable materia respecto de la función Mediana de la variable préstamo

Como se ha dicho, haciendo *clic* en **Cambiar resumen** situado en los cuadros de diálogo de algunos tipos de gráficas (por ejemplo, los de las Figuras 62, 64 y 67) se accede al cuadro de diálogo **Función de resumen** (Figura 90). Aquí se presentan las siguientes alternativas, referidas a la variable o variables (por ejemplo, cuando la estructura de los datos corresponde a **Resúmenes para distintas variables**) declaradas en el recuadro en que **Cambiar resumen** puede tener efecto:

**Media de los valores.** Media aritmética en cada categoría (o de cada variable).

**Mediana de valores.** Mediana en cada categoría (o de cada variable).

**Moda de valores.** Moda en cada categoría (o de cada variable).

**Número de casos.** Número efectivo de casos de cada categoría (o de cada variable).

**Suma de valores.** Suma de los valores en cada categoría (o de cada variable).

**Desviación típica.** Cuasidesviación típica en cada categoría (o de cada variable).

**Varianza.** Cuasivarianza en cada categoría (o de cada variable).

**Valor mínimo.** Mínimo valor en cada categoría (o de cada variable).

**Valor máximo.** Máximo valor en cada categoría (o de cada variable).

**Suma acumulada.** Suma de los valores en cada categoría y en todas las anteriores a ella.

**Porcentaje por encima.** Porcentaje de casos con valores superiores al indicado en el recuadro de **Valor** en cada categoría (o de cada variable).

**Porcentaje por debajo.** Porcentaje de casos con valores inferiores al indicado en el recuadro de **Valor** en cada categoría (o de cada variable).

**Percentiles.** El valor de los datos por debajo del cual se encuentra el porcentaje de valores especificado en **Valor** para cada categoría (o de cada variable).

**Número por encima.** Número de casos con valores superiores al indicado en el recuadro **Valor** en cada categoría (o de cada variable).

**Número por debajo.** Número de casos con valores inferiores al indicado en el recuadro **Valor** en cada categoría (o de cada variable).

**Porcentaje dentro.** Porcentaje de casos (en cada categoría o de cada variable) con valores comprendidos entre los especificados en los recuadros **Menor** y **Mayor**, incluidos ellos mismos.

**Número dentro.** Número de casos (en cada categoría o de cada variable) con valores comprendidos entre los especificados en los recuadros **Menor** y **Mayor**, incluidos ellos mismos.

**Los valores son puntos medios agrupados.** Se obtienen los percentiles considerando que la variable en cuestión representa los puntos medios de los intervalos en que se ha agrupado la variable original.

Por ejemplo, seleccionamos la opción **Gráficos**  $\Rightarrow$  **Barras**. En el cuadro de diálogo resultante (Figura 61) seleccionamos **Simple** y **Resúmenes para grupos de casos**. En el cuadro de diálogo que aparece a continuación (Figura 62) seleccionamos la variable **materia** (disciplina científica del libro) en **Eje de categorías**. En dicho cuadro activamos **Otra función de resumen** y ponemos la variable **préstamo** (número de veces que se ha prestado el libro en el último año) en **Variable**. Pulsamos el botón **Cambiar resumen** y en el cuadro que aparece (Figura 90) seleccionamos **Mediana de valores**. El diagrama de barras resultante es el de la Figura 91.

#### 2.4.7. Cuadro de diálogo de Títulos

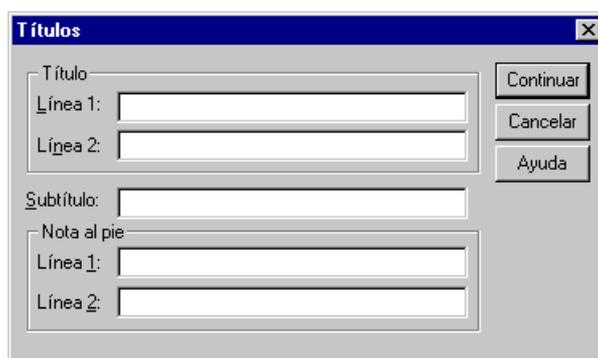


Figura 92: Cuadro de diálogo para colocar Títulos a las representaciones gráficas

Haciendo *click* en el botón **Títulos** del cuadro de diálogo inicial de cada tipo de gráfica (cuadros de diálogo de las Figuras 62, 64, 67 y 69) se muestra un cuadro como el de la Figura 92. Se escriben el título, subtítulo y nota a pie de página a criterio del usuario. El título se puede escribir en dos líneas, así como la nota a pie de página; el subtítulo puede ocupar una sola línea. En cuanto a su extensión, y en cualquier caso, cada línea (recuadro de texto) admite hasta 72 caracteres.

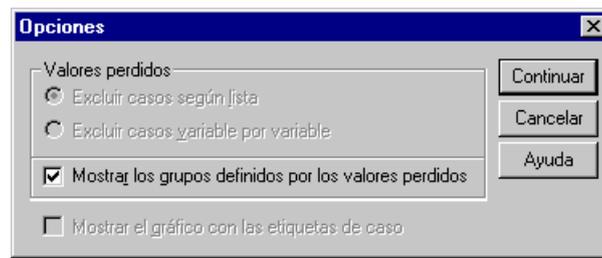


Figura 93: Cuadro de diálogo para seleccionar las Opciones de las representaciones gráficas

#### 2.4.8. Cuadro de diálogo de Opciones

Seleccionando el botón de **Opciones** del cuadro de diálogo inicial de algunos tipos de representaciones gráficas (cuadros de diálogo de las Figuras 62, 64 y 67) nos aparece el cuadro de la Figura 93. En él se controla especialmente el tratamiento de los valores perdidos. Los elementos son:

**Excluir casos según lista.** Se excluyen todos aquellos casos que tengan un valor perdido en alguna de las variables declaradas en la gráfica.

**Excluir casos variable por variable.** Se excluyen para cada variable los casos con valores ausentes en ella.

**Mostrar los grupos definidos por los valores perdidos.** En caso de que haya alguna variable de agrupamiento o un eje de categorías o establecimiento de marcas, etc., se puede representar los grupos de valores perdidos.

**Mostrar el gráfico con las etiquetas de caso.** Esta opción sólo es válida con los diagramas de dispersión. Cada punto del diagrama de dispersión es etiquetado por el valor que le corresponde en la variable insertada en el recuadro **Etiquetar los casos mediante** del cuadro de diálogo inicial de este tipo de gráficas.

#### 2.4.9. Sobre la Plantilla

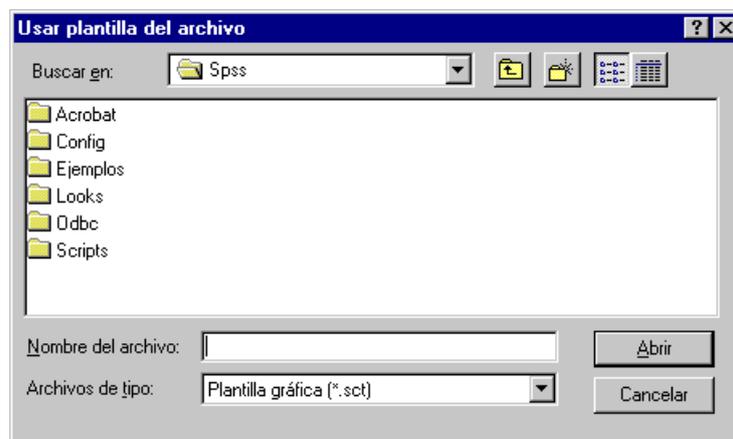


Figura 94: Cuadro de diálogo para usar una Plantilla de un Archivo

El patrón según el cual se crea una determinada gráfica puede ser aplicado a otras de su misma naturaleza o de su misma estructura.

En el cuadro de diálogo inicial de las representaciones gráficas (cuadros de diálogo de las Figuras 62, 64, 67 y 69) se activa la opción **Usar las especificaciones gráficas de** y se pulsa el botón **Archivo**. En el cuadro de diálogo que aparece (Figura 94) se selecciona el subdirectorio donde está el archivo que contiene la plantilla y después el archivo de plantilla deseado (todos tienen extensión *.sct*).

### 3. Probabilidad. Variables aleatorias

#### 3.1. Función de distribución (probabilidad acumulada)

Para obtener el resultado de la función de distribución,  $F(t) = P(X \leq t)$ , con *SPSS* se tiene que utilizar la opción **Transformar**  $\Rightarrow$  **Calcular**, que ya se ha explicado en la sección 1.14. Pero para poder usar esta opción es necesario tener abierto un fichero de datos; debido a lo cual, vamos a crear un nuevo fichero de datos con una variable que tenga un sólo dato.

Por tanto, seleccionamos la opción **Archivo**  $\Rightarrow$  **Nuevo**  $\Rightarrow$  **Datos**. Definimos la primera columna dándole sólo el nombre; que puede ser, por ejemplo, **necesari**. En cuanto a las demás especificaciones de esta variable, dejamos las que tienen el sistema por defecto. Introducimos, en esta primera columna, un sólo dato: por ejemplo, el 1 (podría ser cualquier cosa).

Como los resultados de la función de distribución están comprendidos entre 0 y 1, y van a aparecer en la ventana del editor de datos, podemos especificar como tipo de datos por defecto el *Numérico* con un *Ancho* igual a 8 y con un número de *Decimales* igual a 6. Para ello, seleccionamos **Edición**  $\Rightarrow$  **Opciones**  $\Rightarrow$  *Pestaña de Datos* de la barra de menús. En el apartado **Visualización para nuevas variables numéricas** del cuadro de diálogo correspondiente (Figura 11) ponemos un 8 en **Ancho** y un 6 en **Cifras decimales**.

Recordemos que si seleccionamos **Transformar**  $\Rightarrow$  **Calcular** nos aparece el cuadro de diálogo de la Figura 28. Entre las funciones que aparecen en este cuadro de diálogo, las que calculan los resultados de las funciones de distribución de diversas variables aleatorias para el valor  $c$ ; es decir,  $F(c) = P(X \leq c)$ , son:

**CDF.NORM(valorz)** Calcula la función de distribución de una variable aleatoria normal estándar.

**CDF.BERNOULLI(c,p)** Calcula la función de distribución de una variable aleatoria de Bernoulli con el parámetro de probabilidad **p**.

**CDF.BETA(c,forma1,forma2)** Calcula la función de distribución de una variable aleatoria Beta con los parámetros de forma **forma1** y **forma2**.

**CDF.BINOM(c,n,p)** Calcula la función de distribución de una variable aleatoria binomial (número de éxitos) en los **n** intentos, con una probabilidad de éxito **p**. Cuando **n** es 1, ésta es la misma que la de **CDF.BERNOULLI**.

**CDF.CAUCHY(c,loc,escala)** Calcula la función de distribución de una variable aleatoria de Cauchy con los parámetros **loc** y **escala** dados.

**CDF.CHISQ(c,gl)** Calcula la función de distribución de una variable aleatoria chi-cuadrado con los grados de libertad **gl**.

**CDF.EXP(c,escala)** Calcula la función de distribución de una variable aleatoria exponencial con el parámetro de **escala** dado.

**CDF.F(c,gl1,gl2)** Calcula la función de distribución de una variable aleatoria F de Snedecor con los grados de libertad **gl1** y **gl2**.

**CDF.GAMMA(c,forma,escala)** Calcula la función de distribución de una variable aleatoria Gamma con los parámetros de **forma** y **escala** dados.

**CDF.GEOM(c,p)** Calcula la función de distribución de una variable aleatoria geométrica (número de intentos para obtener un éxito) cuando la probabilidad de éxito es la dada por **p**.

**CDF.HYPER(c,total,muestra,aciertos)** Calcula la función de distribución de una variable aleatoria hipergeométrica (número de objetos con la característica especificada, cuando los objetos de **muestra** son seleccionados aleatoriamente de un universo, de tamaño igual a un **total**, en el que los **aciertos** tienen la característica especificada).

**CDF.LAPLACE(c,media,escala)** Calcula la función de distribución de una variable aleatoria de Laplace con los parámetros de **media** y **escala** especificados.

**CDF.LOGISTIC(c,media,escala)** Calcula la función de distribución de una variable aleatoria logística, con los parámetros de **media** y **escala** especificados.

**CDF.LNORMAL(c,a,b)** Calcula la función de distribución de una variable aleatoria log-normal, con los parámetros **a** y **b** especificados.

**CDF.NEGBIN(c,umbral,p)** Calcula la función de distribución de una variable aleatoria binomial negativa (número de intentos para obtener un número de éxitos igual a **umbral**) siendo la probabilidad de éxito igual a **p**.

**CDF.NORMAL(c,media,desv\_típ)** Calcula la función de distribución de una variable aleatoria normal de parámetros **media** y **desv\_típ**.

- CDF.PARETO(c,umbral,forma)** Calcula la función de distribución de una variable aleatoria de Pareto con los parámetros de **umbral** y **forma** especificados.
- CDF.POISSON(c,media)** Calcula la función de distribución de una variable aleatoria de Poisson con el parámetro de **media** especificado.
- CDF.T(c,gl)** Calcula la función de distribución de una variable aleatoria t de Student con los grados de libertad **gl** especificados.
- CDF.UNIFORM(c,mín,máx)** Calcula la función de distribución de una variable aleatoria uniforme con el **mínimo** y **máximo** especificados.
- CDF.WEIBULL(c,a,b)** Calcula la función de distribución de una variable aleatoria de Weibull con los parámetros **a** y **b** especificados.
- NCDF.BETA(c,forma1,forma2,nc)** Calcula la función de distribución de una variable aleatoria beta no centrada con los parámetros de forma, **forma1** y **forma2**, y no centralidad, **nc**, dados.
- NCDF.CHISQ(c,gl,nc)** Calcula la función de distribución de una variable aleatoria chi-cuadrado no centrada con los grados de libertad **gl** y el parámetro de no centralidad **nc** especificado.
- NCDF.F(c,gl1,gl2,nc)** Calcula la función de distribución de una variable aleatoria F de Snedecor no centrada con los grados de libertad **gl1** y **gl2** y el parámetro de no centralidad **nc** dados.
- NCDF.T(c,gl,nc)** Calcula la función de distribución de una variable aleatoria t de Student no centrada con los grados de libertad **gl** y el parámetro de no centralidad **nc** dados.

### Ejercicios Propuestos

- 3.1.**
- Sea  $X \equiv B(n = 200, p = 0'1)$ . Calcular  $P(X \leq 2)$ ,  $P(X \geq 3)$ ,  $P(X < 5)$ ,  $P(X > 7)$ ,  $P(X = 10)$ .
  - Sea  $X \equiv B(n = 25, p = 0'01)$ . Calcular  $P(X = 3)$ .
  - Sea  $X \equiv B(n = 10, p = 0'77)$ . Calcular  $P(X = 10)$ .
  - Sea  $X \equiv B(n = 12, p = 0'25)$ . Calcular  $P(X = 4)$ .
  - Sea  $X \equiv B(n = 3, p = 0'25)$ . Calcular los resultados de la función de distribución de  $X$  para  $x = 0, 1, 2, 3$ . Deducir los resultados de la función puntual de probabilidad.
  - Sea  $X \equiv B(n = 3, p = 1/6)$ . Calcular los resultados de la función de distribución de  $X$  para  $x = 0, 1, 2, 3$ . Deducir los resultados de la función puntual de probabilidad.
- 3.2.**
- Sea  $X \equiv \mathcal{P}(\lambda = 8)$ . Calcular  $P(X < 6)$ ,  $P(X > 7)$ ,  $P(X \leq 5)$ ,  $P(X \geq 9)$ ,  $P(X = 8)$ .
  - Sea  $X \equiv \mathcal{P}(\lambda = 2)$ . Calcular  $P(X = 0)$ ,  $P(X < 5)$ .
  - Sea  $X \equiv \mathcal{P}(\lambda = 3)$ . Calcular  $P(X = 6)$ .
  - Sea  $X \equiv \mathcal{P}(\lambda = 6)$ . Calcular  $P(X = 3)$ .
  - Sea  $X \equiv \mathcal{P}(\lambda = 9)$ . Calcular  $P(X > 15)$ .
  - Sea  $X \equiv \mathcal{P}(\lambda = 0'6)$ . Calcular  $P(X = 2)$ .
  - Sea  $X \equiv \mathcal{P}(\lambda = 1'8)$ . Calcular  $P(X = 0)$ ,  $P(X = 1)$ ,  $P(X = 2)$ ,  $P(X \geq 2)$ ,  $P(X > 2)$ .
- 3.3.**
- Sea  $X \equiv G(p = 0'25)$ . Calcular  $P(X < 3)$ ,  $P(X > 4)$ ,  $P(X \leq 3)$ ,  $P(X \geq 4)$ ,  $P(X = 5)$ .
  - Sea  $X \equiv G(p = 0'2)$ . Calcular  $P(X = 2)$  y  $P(X > 5)$ .
  - Sea  $X \equiv G(p = 0'6)$ . Calcular  $P(X = 8)$  y  $P(x < 4)$ .
- 3.4.** Sea  $X$  una variable aleatoria normal estándar. Calcular las siguientes probabilidades:
- $P(X \leq -0'34)$ ,  $P(X > 0'76)$ ,  $P(X \leq 1'45)$ ,  $P(X \geq -0'85)$ .
  - $P(0'56 < X < 2'65)$ .
  - $P(|X| > 0'57)$ .
  - $P(|X| < 1'25)$ .
- 3.5.** Sea  $X$  una variable aleatoria normal de media 5 y desviación típica 1'5. Calcular las siguientes probabilidades:
- $P(X \leq 6'5)$ ,  $P(X > 7'5)$ .
  - $P(5'5 < X < 7'5)$ .

c)  $P(|X| > 6'5)$ .

d)  $P(|X| < 7'5)$ .

**3.6.** Sea  $X$  una variable aleatoria que sigue una distribución  $t$  de Student con  $n$  grados de libertad,  $X \equiv t_n$ . Calcular las siguientes probabilidades:

a) Para  $n = 10$ ,  $P(X \leq 1'45)$ ,  $P(X > 0'75)$ .

b) Para  $n = 7$ ,  $P(0'57 < X < 3'5)$ ,  $P(-0'85 < X < -0'25)$ .

c) Para  $n = 15$ ,  $P(|X| > 1'45)$ .

d) Para  $n = 20$ ,  $P(|X| < 3'5)$ .

**3.7.** Sea  $X$  una variable aleatoria que sigue una distribución chi-cuadrado con  $n$  grados de libertad,  $X \equiv \chi_n^2$ . Calcular las siguientes probabilidades:

a) Para  $n = 12$ ,  $P(X > 4'8)$ ,  $P(X \leq 16)$ .

b) Para  $n = 4$ ,  $P(3'3 < X < 9'4)$ .

c) Para  $n = 25$ ,  $P(|X| > 1'5)$ .

d) Para  $n = 14$ ,  $P(|X| < 4'5)$ .

**3.8.** Sea  $X$  una variable aleatoria que sigue una distribución  $F$  de Snedecor con  $m$  y  $n$  grados de libertad,  $X \equiv F_{m,n}$ . Calcular las siguientes probabilidades:

a) Para  $m = 7$ ,  $n = 5$ ,  $P(X > 0'95)$ ,  $P(X \leq 5'6)$ .

b) Para  $m = 15$ ,  $n = 10$ ,  $P(0'4 < X < 2'4)$ .

c) Para  $m = 12$ ,  $n = 14$ ,  $P(|X| < 6'3)$ .

d) Para  $m = 10$ ,  $n = 8$ ,  $P(|X| > 3'5)$ .

## 3.2. Inversa de la función de distribución (percentiles)

Para obtener el resultado de la inversa de la función de distribución (los percentiles) con *SPSS* se tiene que utilizar la opción **Transformar**  $\Rightarrow$  **Calcular**. Igual que antes, para poder usar esta opción es necesario tener abierto un fichero de datos.

Como los resultados de los percentiles pueden ser cualesquiera, y van a aparecer en la ventana del editor de datos, podemos especificar como tipo de datos por defecto el *Numérico* con un *Ancho* igual a 8 y con un número de *Decimales* igual a 4. Para ello, seleccionamos **Edición**  $\Rightarrow$  **Opciones**  $\Rightarrow$  *Pestaña de Datos* de la barra de menús. En el apartado **Visualización para nuevas variables numéricas** del cuadro de diálogo correspondiente (Figura 11) ponemos un 8 en *Ancho* y un 4 en *Cifras decimales*.

Entre las funciones que aparecen en el cuadro de diálogo de la opción **Transformar**  $\Rightarrow$  **Calcular** (Figura 28) las que calculan los resultados de los percentiles (valores  $a$  tales que  $P(X \leq a) = p$ ) son:

**IDF.BETA(p,forma1,forma2)** Calcula el valor de la distribución Beta, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.CAUCHY(p,loc,escala)** Calcula el valor de la distribución de Cauchy, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.CHISQ(p,gl)** Calcula el valor de la distribución de chi-cuadrado, con los grados de libertad **gl** especificados, para el cual la probabilidad acumulada es **p**.

**IDF.EXP(p,escala)** Calcula el valor de una variable exponencial, con una tasa de decrecimiento **escala**, para el cual la probabilidad acumulada es **p**.

**IDF.F(p,gl1,gl2)** Calcula el valor de la distribución F de Snedecor, con los grados de libertad especificados, para el cual la probabilidad acumulada es **p**.

**IDF.GAMMA(p,forma,escala)** Calcula el valor de la distribución Gamma, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.LAPLACE(p,media,escala)** Calcula el valor de la distribución de Laplace, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.LOGISTIC(p,media,escala)** Calcula el valor de la distribución logística, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.LNORMAL(p,a,b)** Calcula el valor de la distribución log-normal, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.NORMAL(p,media,desv\_típ)** Calcula el valor de la distribución normal, con la media y la desviación típica especificadas, para el cual la probabilidad acumulada es **p**.

**IDF.PARETO(p,umbral,forma)** Calcula el valor de la distribución de Pareto, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.T(p,gl)** Calcula el valor de la distribución *t* de Student, con los grados de libertad **gl** especificados, para el cual la probabilidad acumulada es **p**.

**IDF.UNIFORM(p,mín,máx)** Calcula el valor de la distribución uniforme, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**IDF.WEIBULL(p,a,b)** Calcula el valor de la distribución de Weibull, con los parámetros especificados, para el cual la probabilidad acumulada es **p**.

**PROBIT(p)** Calcula el valor de una distribución normal típica con una probabilidad acumulada igual a **p**.

### Ejercicios Propuestos

3.9. Hallar el valor de los siguientes percentiles:

$$Z_{0'58}, \quad Z_{0'42}, \quad Z_{0'999}, \quad Z_{0'001}.$$

3.10. Si  $Z$  denota la variable aleatoria Normal Estándar, calcular el valor de  $a$  para que se verifique:

a)  $P(Z \leq a) = 0'72$ .

b)  $P(Z < a) = 0'34$ .

c)  $P(Z > a) = 0'28$ .

d)  $P(Z \geq a) = 0'97$ .

3.11. Si  $X$  es una variable aleatoria con distribución Normal de media  $3'5$  y desviación típica  $0'8$ , determinar el valor de  $a$  tal que:

a)  $P(X \leq a) = 0'77$ .

b)  $P(X < a) = 0'012$ .

c)  $P(X > a) = 0'06$ .

d)  $P(X \geq a) = 0'99$ .

3.12. Calcular el valor de los siguientes percentiles:

a)  $\chi_{6, 0'01}^2$ .

b)  $\chi_{6, 0'99}^2$ .

c)  $\chi_{72, 0'975}^2$ .

3.13. Sea  $X$  una variable aleatoria que sigue una distribución chi-cuadrado de Pearson con 15 grados de libertad. Determinar el valor de  $a$  que verifica la siguiente igualdad:

a)  $P(X \leq a) = 0'05$ .

b)  $P(X > a) = 0'99$ .

c)  $P(|X| < a) = 0'5$ .

3.14. Calcular el valor de los siguientes percentiles:

a)  $t_{26, 0'9}$ .

b)  $t_{26, 0'1}$ .

c)  $t_{75, 0'8}$ .

3.15. Sea  $X$  una variable aleatoria que sigue una distribución *t* de Student con 20 grados de libertad. Determinar el valor de  $a$  que verifica la siguiente igualdad:

**a)**  $P(X \leq a) = 0'99$ .

**b)**  $P(X \geq a) = 0'25$ .

**c)**  $P(|X| < a) = 0'9$ .

**3.16.** Calcular el valor de los siguientes percentiles:

**a)**  $F_{8, 6, 0'975}$ .

**b)**  $F_{25, 50, 0'01}$ .

**c)**  $F_{45, 35, 0'01}$ .

**3.17.** Sea  $X$  una variable aleatoria que sigue una distribución  $F$  de Snedecor con 10 grados de libertad en el numerador y 8 grados de libertad en el denominador. Determinar el valor de  $a$  que verifica la siguiente igualdad:

**a)**  $P(X < a) = 0'9$ .

**b)**  $P(X > a) = 0'05$ .

**c)**  $P(|X| < a) = 0'98$ .

## 4. Contrastes paramétricos

### 4.1. Contraste para una media. Intervalo de confianza



Figura 95: Cuadro de diálogo para hacer un test de hipótesis sobre una media

Para realizar un contraste de hipótesis sobre una media se ha de seleccionar **Anализar** ⇒ **Comparar medias** ⇒ **Prueba T para una muestra**. Este procedimiento sólo realiza el contraste bilateral de la hipótesis nula  $H_0 : \mu = \mu_0$  frente a la hipótesis alternativa  $H_1 : \mu \neq \mu_0$  en el caso en el que la varianza poblacional  $\sigma^2$  es desconocida. Recordemos que para aplicar esta técnica (test de la  $t$  de Student para una muestra) se ha de cumplir que la variable sea Normal o que el tamaño muestral sea grande (mayor o igual que 30).

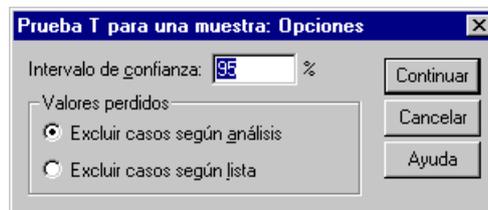


Figura 96: Opciones de un test de hipótesis sobre una media poblacional

En el recuadro **Contrastar Variables** del cuadro de diálogo que aparece (Figura 95) se selecciona la variable o variables para las cuales se va a realizar este tipo de contraste. Y en el recuadro **Valor de prueba** se especifica el valor,  $\mu_0$ , con el que se compara la media poblacional.

Si pulsamos el botón **Opciones** de dicho cuadro de diálogo nos aparece el cuadro de la Figura 96. Estas opciones son:

**Intervalo de confianza.** Por defecto se muestra un intervalo de confianza al 95 % para la diferencia entre la media poblacional,  $\mu$ , y el valor de contraste hipotético,  $\mu_0$ . Se puede introducir un valor entre 1 y 99 para solicitar otro nivel de confianza.

**Valores perdidos.** Si se desea contrastar varias variables y existen datos perdidos en una o en varias de las variables, se puede indicar qué casos deberán ser excluidos.

**Excluir casos según análisis.** Cada contraste utiliza todos los casos que tienen datos válidos para la variable contrastada. Los tamaños muestrales pueden variar de una prueba a otra.

**Excluir casos según lista.** Cada contraste utiliza únicamente los casos que tienen datos válidos para todas las variables utilizadas en cualquiera de los contrastes solicitados. El tamaño muestral es constante en todas las pruebas.

	N	Media	Desviación típ.	Error típ. de la media
PRÉSTAMO	30	13,60	9,74	1,78

Figura 97: Estadísticos en el contraste sobre la media de la variable préstamo

Como el tamaño muestral de nuestro fichero **Libros.sav** es igual a 30, se puede aplicar este procedimiento a cualquier variable. Dado que la media muestral de la variable **préstamo** es igual a 13'6, vamos a contrastar la hipótesis nula  $H_0 : \mu = 14$  frente a la hipótesis alternativa  $H_1 : \mu \neq 14$ . Los resultados que aparecen en el visor de **SPSS** son los de las Figuras 97 y 98.

Como el nivel crítico, *Sig. (bilateral)*, es igual a 0'824 (claramente mayor que el nivel de significación  $\alpha = 0'05$ ) entonces aceptamos la hipótesis nula; es decir, la media poblacional de la variable **préstamo** puede considerarse igual a 14. También obtenemos el resultado del estadístico de contraste:

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = -0'225,$$

	Valor de prueba = 14					
	t	gl	Sig. (bilateral)	Diferencia de medias	Intervalo de confianza para la diferencia	
					Inferior	Superior
PRÉSTAMO	-,225	29	,824	-,40	-4,04	3,24

Figura 98: Resultados del contraste sobre la media de la variable préstamo

los grados de libertad de la  $t$  de Student:

$$gl = 29,$$

la diferencia entre la media muestral y el valor con el se compara:

$$\text{Diferencia de medias} = \bar{x} - \mu_0 = -0'4,$$

y el intervalo de confianza al 95 % para la diferencia entre la media poblacional,  $\mu$ , y la media hipotética, 14:

$$\text{Intervalo de confianza para la diferencia} = (-4'04, 3'24).$$

De esto último podemos obtener un intervalo de confianza al 95 % para la media poblacional:

$$\begin{aligned} -4'04 < \mu - 14 < 3'24 &\Rightarrow \\ -4'04 + 14 < \mu < 3'24 + 14 &\Rightarrow \\ 9'96 < \mu < 17'24. & \end{aligned}$$

## 4.2. Comparación de dos medias con muestras independientes

Para realizar un contraste de hipótesis sobre comparación de dos medias con muestras independientes se tiene que seleccionar **Analizar**  $\Rightarrow$  **Comparar medias**  $\Rightarrow$  **Prueba T para muestras independientes**. Aparece entonces el cuadro de diálogo de la Figura 99.



Figura 99: Cuadro de diálogo para hacer un test de hipótesis sobre comparación de dos medias con muestras independientes

Este procedimiento realiza el contraste bilateral de la hipótesis nula  $H_0 : \mu_1 = \mu_2$  frente a la hipótesis alternativa  $H_1 : \mu_1 \neq \mu_2$  en el caso de muestras independientes y varianzas poblacionales desconocidas, distinguiendo los casos en que éstas sean iguales o distintas. Por ello, también presenta un contraste de comparación de varianzas poblacionales, que es la Prueba de Levene (un contraste más independiente que otros respecto de la hipótesis de normalidad de las variables implicadas). Recordemos que para aplicar esta técnica (test de la  $t$  de Student para dos muestras) se ha de cumplir que las variables sean Normales o que los tamaños muestrales sean grandes (mayores o iguales que 30).

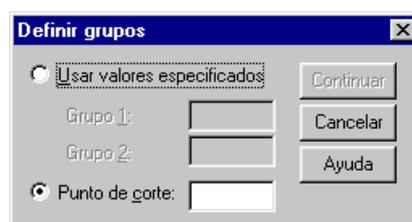


Figura 100: Cuadro de diálogo para definir los grupos en una comparación de dos medias con muestras independientes

En **Contrastar variables** se especifica la variable (o variables) objeto de estudio. Téngase presente que si se especifica más de una variable, éstas serán analizadas individualmente. En el recuadro **Variable de agrupación** se ha de especificar una variable dicotómica o que se vaya a dicotomizar, pues queremos comparar las medias de la variable elegida en **Contrastar variables** para los dos grupos definidos por las dos categorías de la variable especificada en **Variable de agrupación**. A continuación se ha de pulsar el botón **Definir grupos** para especificar los dos grupos en los que se va a dividir esta última variable (ver Figura 100).

Los dos grupos se pueden crear basándose en una variable numérica o no numérica. Si la variable especificada en **Variable de agrupación** es numérica (o numéricamente codificada) se puede dar, a su vez, dos casos: en el primero la variable es ya dicotómica, por lo que se activa **Usar valores especificados** (Figura 100) y se designan los dos grupos (**Grupo 1** y **Grupo2**) según los valores originales. Si la variable numérica no es dicotómica, el recuadro **Punto de corte** tiene como objetivo el especificar el valor de la variable que dará lugar a una partición en dos categorías, una de las cuales será el conjunto de los valores menores que el especificado, y la otra el resto de valores.

Si la variable especificada en **Variable de agrupación** no es numérica, el cuadro de diálogo es como el anterior (Figura 100) si exceptuamos **Punto de corte**. Si la variable es ya dicotómica se designan los dos grupos (**Grupo 1** y **Grupo2**) según los valores originales. Si la variable no es dicotómica, se especifican sólo dos de las categorías de dicha variable, quedando excluidas del análisis el resto de ellas.

Igual que en el caso del contraste sobre una media, en el cuadro de diálogo de la Figura 99 también tenemos la posibilidad de activar **Opciones**, cuya interpretación es completamente análoga a la del contraste mencionado (ver Figura 96).

	VEJEZ	N	Media	Desviación típ.	Error típ. de la media
PRECIO	>= 3	5	5680,00	3298,22	1475,01
	< 3	25	8507,20	6365,79	1273,16

Figura 101: Estadísticos en el contraste sobre comparación de dos medias con muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ de la diferencia	Intervalo de confianza para la diferencia	
									Inferior	Superior
PRECIO	Se han asumido varianzas iguales	2,215	,148	-,958	28	,346	-2827,20	2951,13	-8872	3217,92
	No se han asumido varianzas iguales			-1,45	11,149	,174	-2827,20	1948,48	-7109	1454,39

Figura 102: Resultados del contraste sobre comparación de dos medias con muestras independientes

Con los datos de nuestro fichero **Libros.sav**, supongamos que queremos saber si la media del precio de los libros antiguos es igual a la media del precio de los libros recientes. Recordemos que los resultados de la variable **vejez** significan: 1= de 1 a 3 años transcurridos desde su publicación, 2= de 4 a 6 años transcurridos desde su publicación, 3= de 7 a 9 años transcurridos desde su publicación, 4= 10 o más años transcurridos desde su publicación. Podemos dicotomizar la variable **vejez** considerando que los libros recientes son los que tienen como resultado 1 o 2, y los libros antiguos son los que tienen como resultado 3 o 4. Por tanto, en **Contrastar variables** colocamos **precio** y en **Variable de agrupación** seleccionamos **vejez**. Pulsamos, a continuación, el botón **Definir grupos**, y en el cuadro que resulta (Figura 100) activamos la opción **Punto de corte** y escribimos en el recuadro el valor 3 (un grupo está formado por los valores menores que 3, y el otro por el resto). En el visor de **SPSS** aparecen los resultados que se muestran en las Figuras 101 y 102

El tamaño muestral del grupo 1 (resultado de **vejez** mayor o igual que 3) es sólo 5, y el tamaño muestral del grupo 2 (resultado de **vejez** menor que 3) es 25. Ninguno de estos tamaños son grandes (mayores o iguales que 30), pero suponemos que la variable **precio** es Normal (consultar el apartado 5.5 para comprobar que es cierto). Por tanto, se puede utilizar este procedimiento de comparación de dos medias poblacionales. Para cada uno de los dos grupos, la tabla de la Figura 101 nos da los resultados de las medias muestrales (*Media*), las cuasidesviaciones típicas muestrales (*Desviación típ.*) y los errores típicos de las medias (*Error típ. de la media*). El error típico de la media es la desviación típica del estadístico media muestral, que es igual al valor de la cuasidesviación típica dividido por la raíz cuadrada del tamaño muestral.

En cuanto al resultado del contraste de hipótesis sobre igualdad de varianzas (*Prueba de Levene*) el valor del estadístico de contraste es 2'215 y el nivel crítico (*Sig.*) es 0'148. Como este nivel crítico es claramente mayor que 0'05 entonces se puede aceptar la hipótesis de igualdad de varianzas poblacionales. Tenemos que ver, por tanto, los resultados del contraste de comparación de medias que hay a continuación de *Se han asumido varianzas iguales*. El estadístico de contraste para la

comparación de medias con varianzas desconocidas e iguales es:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = -0'958,$$

los grados de libertad son:

$$gl = 28,$$

el nivel crítico es:

$$\text{Sig. (bilateral)} = 0'346.$$

Como dicho nivel crítico es claramente mayor que 0'05, debemos aceptar la hipótesis nula; es decir, la media del precio de los libros recientes es igual a la media del precio de los libros antiguos (en las poblaciones). Además aparece el resultado de la diferencia entre las dos medias muestrales,  $\bar{x}_1 - \bar{x}_2$  (que es igual al numerador del estadístico de contraste):

$$\text{Diferencia de medias} = -2827'2,$$

el error típico de la diferencia (que es igual al denominador del estadístico de contraste):

$$2951'13,$$

y un intervalo de confianza al 95 % para la diferencia de medias poblacionales:

$$-8872'32 < \mu_1 - \mu_2 < 3217'92.$$

#### 4.2.1. Ejemplo de edición de unos datos para hacer la comparación de dos medias con muestras independientes

Se está realizando el estudio de una base de datos e interesa saber si los resúmenes que publica escritos en francés son significativamente diferentes en longitud (número de palabras) a los escritos en inglés. La muestra estudiada se ha cogido al azar y consiste en 30 resúmenes escritos en francés y 30 (independientes de los otros) en inglés.

francés	70	65	68	74	79	67	75	80	62	69
	61	57	71	74	82	91	70	64	72	67
	74	70	81	85	70	74	75	71	69	54
inglés	80	47	59	67	89	57	72	78	74	72
	104	118	89	87	79	78	101	120	107	95
	85	87	90	98	89	75	90	101	85	94

Hay que realizar un contraste de comparación de dos medias con muestras independientes. Pero, en primer lugar tenemos que introducir estos datos en el editor y luego grabarlos en un nuevo fichero. La primera columna (que llamaremos **longitud**) va a contener todos los resultados de las dos muestras, una debajo de la otra. Y la segunda columna de datos (que llamaremos **grupo**) va a tener un 1 (para los resultados que provengan de la muestra de resúmenes escritos en francés) o un 2 (para los resultados que provengan de la muestra de resúmenes escritos en inglés).

Para introducir estos datos en el editor de datos de **SPSS**, en primer lugar seleccionamos **Archivo** ⇒ **Nuevo** ⇒ **Datos**. En segundo lugar definimos convenientemente las dos variables. A continuación introducimos los datos de la manera indicada (ver la Figura 103. Y por último grabamos estos datos en un nuevo fichero llamado **LongitudResumenes.sav**

Ahora ya podemos, con los datos del fichero **LongitudResumenes.sav**, realizar el contraste mencionado. Para ello, en el cuadro de diálogo de la Figura 99 se selecciona la variable **longitud** en el recuadro **Contrastar variables**; y la variable **grupo** en el recuadro **Variable de agrupación**. En el cuadro de diálogo de la Figura 100 se activa **Usar valores especificados**, y se coloca un **1** junto a **Grupo 1** y un **2** junto a **Grupo 2**. Los resultados de este contraste se muestran en la tabla de la Figura 104.

Observamos que, en el contraste de comparación de varianzas poblacionales, el nivel crítico (*Sig.*) es igual a 0'003 (menor que 0'05) por lo que rechazamos la hipótesis de igualdad de varianzas, y por tanto aceptamos que las varianzas poblacionales son distintas. Observamos que el nivel crítico (*Sig. (bilateral)*) para el contraste de comparación de medias con varianzas desconocidas y distintas es 0 (menor que 0'05) por lo que rechazamos la hipótesis de igualdad de medias. Por tanto, aceptamos que la media del número de palabras por resumen escrito en francés es distinta de la media del número de palabras por resumen escrito en inglés. Además, como en las muestras se obtiene que la media (muestral) de los de francés es menor que la media (muestral) de los de inglés, entonces debe aceptarse que ocurre lo mismo en las poblaciones respectivas.

	longitud	grupo
1	70	1
2	65	1
3	68	1
4	74	1
5	79	1
6	67	1
7	75	1
8	80	1
9	62	1
10	69	1
11	61	1
12	57	1
13	71	1
14	74	1
15	82	1

	longitud	grupo
16	91	1
17	70	1
18	64	1
19	72	1
20	67	1
21	74	1
22	70	1
23	81	1
24	85	1
25	70	1
26	74	1
27	75	1
28	71	1
29	69	1
30	54	1

	longitud	grupo
31	80	2
32	47	2
33	59	2
34	67	2
35	89	2
36	57	2
37	72	2
38	78	2
39	74	2
40	72	2
41	104	2
42	118	2
43	89	2
44	87	2
45	79	2

	longitud	grupo
46	78	2
47	101	2
48	120	2
49	107	2
50	95	2
51	85	2
52	87	2
53	90	2
54	98	2
55	89	2
56	75	2
57	90	2
58	101	2
59	85	2
60	94	2

Figura 103: Datos del fichero LongitudResumenes.sav

#### Estadísticos del grupo

	GRUPO	N	Media	Desviación típ.	Error típ. de la media
LONGITUD	1	30	71,37	7,95	1,45
	2	30	85,57	16,67	3,04

#### Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ de la diferencia	Intervalo de confianza para la diferencia	
									Inferior	Superior
LONGITUD	Se han asumido varianzas iguales	9,884	,003	-4,211	58	,000	-14,20	3,37	-20,95	-7,45
	No se han asumido varianzas iguales			-4,211	41,52	,000	-14,20	3,37	-21,01	-7,39

Figura 104: Resultados del contraste sobre comparación de las medias de las longitudes de los resúmenes

### 4.3. Comparación de dos medias con muestras relacionadas

Para realizar un contraste de hipótesis sobre comparación de dos medias con muestras relacionadas se tiene que seleccionar **Analizar** ⇒ **Comparar medias** ⇒ **Prueba T para muestras relacionadas**. Aparece entonces el cuadro de diálogo de la Figura 106.

Este procedimiento realiza el contraste bilateral de la hipótesis nula  $H_0 : \mu_1 = \mu_2$  frente a la hipótesis alternativa  $H_1 : \mu_1 \neq \mu_2$  en el caso de muestras apareadas. Recordemos que para aplicar esta técnica (test de la  $t$  de Student para dos muestra) se ha de cumplir que las variables sean Normales o que el tamaño muestral sea grande (mayor o igual que 30).

Hay que declarar un par de variables en el recuadro **Variables relacionadas**. Para ello debemos seleccionar (marcar) un par de variables de la lista; veremos como éstas quedarán escritas automáticamente junto a **Variable 1** y **Variable 2** del grupo **Selecciones actuales**. Después pulsamos el botón .

Igual que en el caso del contraste anterior, en el cuadro de diálogo de la Figura 106 también tenemos la posibilidad de activar **Opciones**, cuya interpretación es completamente análoga a la del contraste mencionado (ver Figura 96).

Como el tamaño muestral es 30, podemos aplicar este contraste a las variables **opinión1** y **opinión2**. Los resultados obtenidos se muestran en la Figura 105.

**Estadísticos de muestras relacionadas**

	Media	N	Desviación típ.	Error típ. de la media
Par 1 OPINIÓN1	2,93	30	1,41	,26
OPINIÓN2	3,37	30	1,40	,26

**Correlaciones de muestras relacionadas**

	N	Correlación	Sig.
Par 1 OPINIÓN1 y OPINIÓN2	30	-,092	,630

**Prueba de muestras relacionadas**

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par 1 OPINIÓN1 - OPINIÓN2	-,43	2,08	,38	-1,21	,34	-1,142	29	,263

Figura 105: Resultados del contraste sobre comparación de dos medias con muestras relacionadas



Figura 106: Cuadro de diálogo para hacer un test de hipótesis sobre comparación de dos medias con muestras relacionadas

En la primera tabla aparecen los resultados de los estadísticos para cada una de las variables. Las medias muestrales son  $\bar{x}_1 = 2'93$  y  $\bar{x}_2 = 3'37$ ; las cuasidesviaciones típicas muestrales son  $S_1 = 1'41$  y  $1'4$ ; los errores típicos de las medias muestrales son  $S_1/\sqrt{n} = 0'26$  y  $S_2/\sqrt{n} = 0'26$ , respectivamente.

En la segunda tabla aparecen los resultados de la correlación entre las dos variables. El valor del coeficiente de correlación lineal de Pearson en la muestra es  $r = -0'092$ . También aparece en esta tabla el resultado del nivel crítico para la hipótesis nula de igualdad a cero del coeficiente de correlación lineal poblacional. Este nivel crítico (Sig.) es igual a  $0'630$ , claramente mayor que  $0'05$ , por lo que podemos aceptar la hipótesis nula. Por tanto, aceptamos que las dos variables son independientes en la población de todos los libros.

En la tercera tabla aparecen los siguientes resultados: la media de la variable diferencia  $D = X_1 - X_2$  es igual a:

$$\text{Media} = \bar{d} = -0'43,$$

la cuasidesviación típica de la variable diferencia  $D$  es igual a:

$$\text{Desviación típ.} = S_d = 2'08,$$

la desviación típica del estadístico  $\bar{d}$  es igual a:

$$\text{Error típ. de la media} = \frac{S_d}{\sqrt{n}} = 0'38,$$

un intervalo de confianza para la diferencia de las medias poblacionales es:

$$-1'21 < \mu_1 - \mu_2 < 0'34,$$

el estadístico de contraste para la comparación de las dos medias poblacionales (con muestras relacionadas) es:

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = -1'142,$$

los grados de libertad son:

$$gl = 29,$$

y el nivel crítico es:

$$\text{Sig. (bilateral)} = 0'263.$$

Como dicho nivel crítico es claramente mayor que 0'05 se puede aceptar la hipótesis nula de la igualdad de las medias poblacionales de las dos variables. Por tanto, aceptamos que son iguales las medias de las opiniones de los dos primeros expertos en encuadernación.

## 5. Contrastes no paramétricos

### 5.1. Contraste sobre una proporción

Para hacer un contraste de hipótesis sobre una proporción poblacional se selecciona **Analizar**  $\Rightarrow$  **Pruebas no paramétricas**  $\Rightarrow$  **Prueba Binomial**. Si la variable estudiada no es dicotómica, debe dicotomizarse como veremos a continuación. Se especifican,

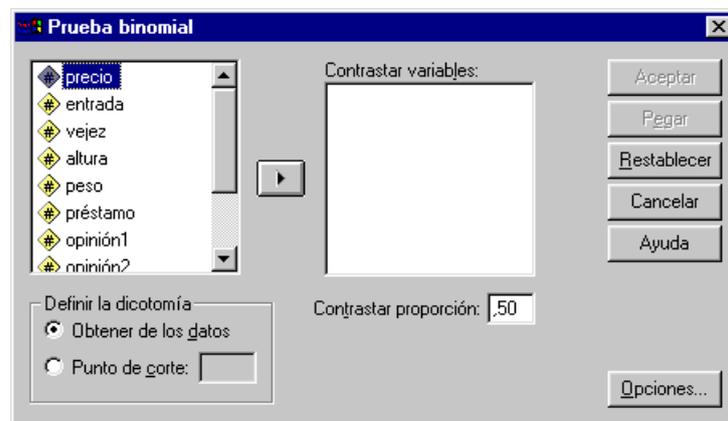


Figura 107: Cuadro de diálogo para el contraste sobre una proporción

como de costumbre, las variables a tratar en **Contrastar variables** teniendo presente que no tienen por qué ser originariamente dicotómicas (Figura 107). Dada la naturaleza del contraste, se examinan las variables de una en una. Si se hace *click* en **Aceptar**, las variables han de ser dicotómicas y, por defecto, se obtienen los resultados del contraste en el caso en que la proporción poblacional esperada sea igual a 0'5.

Si la variable no es dicotómica, en **Definir la dicotomía** se activa la opción **Punto de corte** y en el recuadro adjunto se escribe el valor de corte, de forma que los inferiores o iguales a él se agruparán en la primera categoría y el resto en la otra. Si no queremos que las proporciones de ambas categorías sean iguales (0'5), se puede especificar en el recuadro **Contrastar proporción** la proporción (o probabilidad) esperada poblacional de la primera categoría.

Supongamos que queremos contrastar la hipótesis nula de que la proporción poblacional de libros que se prestan 15 veces o menos cada año es igual a 0'4. Para ello, en **Contrastar variables** elegimos la variable **préstamo**; activamos la opción **Punto de corte** y en el recuadro adjunto escribimos 15; en **Contrastar proporción** escribimos ,4 (se ha de omitir el cero que hay delante de la coma decimal) y luego pulsamos **Aceptar**. Los resultados aparecen en la Figura 108. Como vemos, junto al nivel crítico (*Sig. asintót.*) aparece la palabra *unilateral*; esto significa que la hipótesis alternativa que se ha considerado es que la proporción poblacional sea mayor que 0'4 (recordemos que la hipótesis alternativa debe ser compatible con los resultados de la muestra). Como el nivel crítico es 0'003 (claramente menor que 0'05) entonces se rechaza la hipótesis nula y por tanto se acepta la alternativa; es decir, se acepta que la proporción poblacional de libros que se prestan anualmente 15 veces o menos es mayor que 0'4.

	Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (unilateral)
PRÉSTAMO	Grupo 1	<= 15	20	,7	,003 <sup>a</sup>
	Grupo 2	> 15	10	,3	
	Total		30	1,0	

a. Basado en la aproximación Z.

Figura 108: Contraste sobre una proporción para la variable préstamo

### 5.2. Contraste $\chi^2$ de Pearson sobre proporciones

Para realizar este contraste se selecciona **Analizar**  $\Rightarrow$  **Pruebas no paramétricas**  $\Rightarrow$  **Chi cuadrado**. Este procedimiento tabula una variable en categorías y calcula un estadístico chi-cuadrado. Esta prueba de bondad de ajuste compara las frecuencias observadas y esperadas en cada categoría para contrastar si todas las categorías contienen la misma proporción de valores o si cada categoría contiene una proporción de valores especificada por el usuario.

En el recuadro **Contrastar variables** del cuadro de diálogo inicial (Figura 109) se seleccionan las variables objeto de estudio. Haciendo *click* en **Aceptar** se considera, por defecto, que cada categoría de la variable tiene la misma frecuencia esperada.

Con el fin de especificar el rango de valores que serán objeto de análisis, en **Rango esperado** se presentan dos alternativas:

**Obtener de los datos.** Por defecto, cada valor de la variable es considerado una categoría.

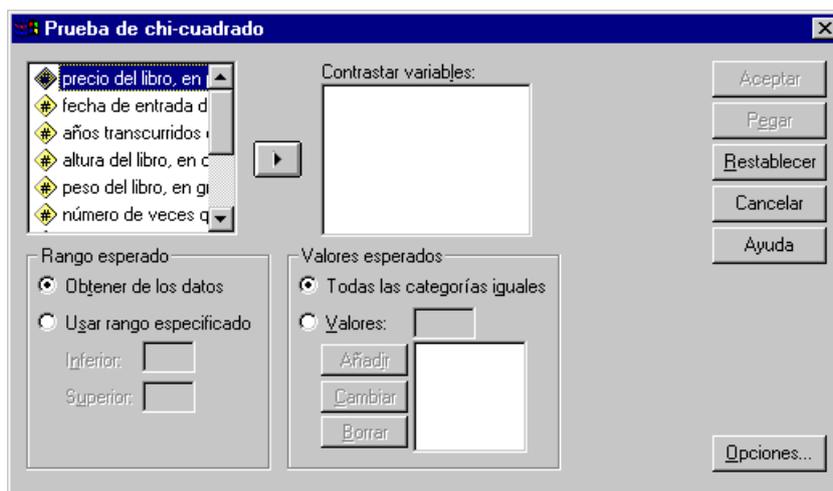


Figura 109: Cuadro de diálogo para el contraste chi-cuadrado sobre proporciones

**Usar rango especificado.** En caso de que se quiera analizar sólo un determinado rango de valores, se especifica en los recuadros de texto **Inferior** y **Superior** los valores, enteros, entre los que están aquellos que interesan al usuario, incluidos ellos mismos.

En el grupo **Valores esperados** tenemos dos opciones:

**Todas las categorías iguales.** Por defecto, todas las categorías tienen la misma frecuencia esperada (misma proporción).

**Valores.** Si no fuese éste el caso, el usuario tiene la posibilidad de especificar en el recuadro de texto adjunto las frecuencias absolutas esperadas. También se pueden especificar porcentajes o frecuencias relativas (proporciones). Hay que ser cuidadoso con el orden en que se especifican las frecuencias, pues corresponden a los valores de la variable en orden ascendente. Si introducimos frecuencias esperadas, la suma de todas ellas debe ser igual al tamaño muestral; si son frecuencias relativas o proporciones, la suma de todas ellas debe ser igual a 1 (en este caso la separación decimal se hace con una coma); y si son porcentajes, la suma debe ser igual a 100.

Por ejemplo, en **Contrastar variables** seleccionamos la variable **vejez** de nuestro fichero de datos **Libros.sav**. Recordemos que esta variable tiene 4 resultados distintos: 1, 2, 3 y 4. Supongamos que queremos contrastar la hipótesis nula de que las proporciones de estos resultados, en la población, son 0'4, 0'3, 0'2 y 0'1, respectivamente; es decir:

$$P(X = 1) = 0'4, \quad P(X = 2) = 0'3, \quad P(X = 3) = 0'2, \quad P(X = 4) = 0'1.$$

En el grupo **Rango esperado** debemos activar **Obtener de los datos**. En el grupo **Valores esperados** debemos seleccionar **Valores**. A continuación debemos especificar las frecuencias (o los porcentajes o las proporciones) esperadas. Si queremos especificar las proporciones (frecuencias relativas) esperadas, hacemos lo siguiente:

- En **Valores** escribimos 0, 4 y pulsamos **Añadir**. Esto significa que la proporción (o frecuencia relativa) esperada para el primer resultado de la variable (que es 1) es igual a 0, 4.
- En **Valores** escribimos 0, 3 y pulsamos **Añadir**. Esto significa que la proporción (o frecuencia relativa) esperada para el segundo resultado de la variable (que es 2) es igual a 0, 3.
- En **Valores** escribimos 0, 2 y pulsamos **Añadir**. Esto significa que la proporción (o frecuencia relativa) esperada para el tercer resultado de la variable (que es 3) es igual a 0, 2.
- En **Valores** escribimos 0, 1 y pulsamos **Añadir**. Esto significa que la proporción (o frecuencia relativa) esperada para el cuarto resultado de la variable (que es 4) es igual a 0, 1.

Si se quiere especificar los porcentajes esperados, hacemos lo mismo pero en vez de escribir 0, 4, 0, 3, 0, 2 y 0, 1 se escribe 40, 30, 20 y 10. Por último, si se quiere especificar las frecuencias absolutas esperadas, tecleamos los siguientes valores: 12, 9, 6 y 3 (pues son el 40 % de 30, el 30 % de 30, el 20 % de 30 y el 10 % de 30, respectivamente). Los resultados se muestran en la Figura 110. Como el nivel crítico (*Sig. asintót.*) es igual a 0'317 (claramente mayor que 0'05) aceptamos la hipótesis nula. Pero esa conclusión no es muy fiable porque para aplicar el contraste  $\chi^2$  de Pearson de bondad de ajuste es conveniente que haya como máximo un 20 % de frecuencias esperadas menores que 5; sin embargo en este ejemplo hay un 25 % de frecuencias esperadas menores que 5.

VEJEZ

	N observado	N esperado	Residual
1	15	12,0	3,0
2	10	9,0	1,0
3	2	6,0	-4,0
4	3	3,0	,0
Total	30		

Estadísticos de contraste

	VEJEZ
Chi-cuadrado <sup>a</sup>	3,528
gl	3
Sig. asintót.	,317

a. 1 casillas (25,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 3,0.

Figura 110: Resultados del contraste chi-cuadrado sobre proporciones para la variable vejez

### 5.3. Contraste $\chi^2$ sobre independencia de dos variables

Para obtener los resultados del contraste  $\chi^2$  de Pearson sobre independencia de variables aleatorias se selecciona **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Tablas de contingencia**.

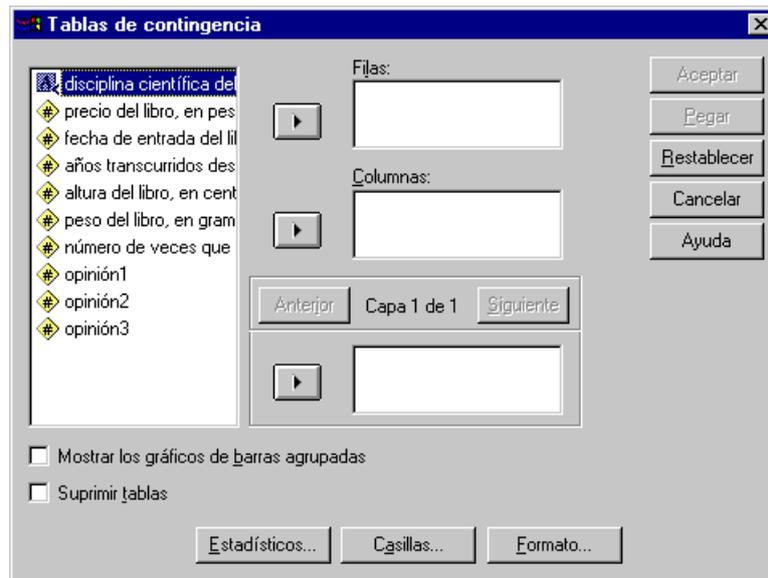


Figura 111: Cuadro de diálogo de la opción Tablas de contingencia

Como se ve en la Figura 111, podemos seleccionar las variables que constituyen las **Filas** por un lado, y las **Columnas** por otro. El que se especifique más de una variable de cada tipo no implica que se vayan a tratar todas conjuntamente, sino que se procesan las posibles combinaciones de dos en dos. Por defecto, haciendo *click* en **Aceptar**, y sin más especificaciones, se obtienen las tablas de contingencia (o de doble entrada) con las frecuencias absolutas conjuntas y marginales. Por ejemplo, si en **Filas** seleccionamos la variable **opinión1** y en **Columnas** la variable **opinión2** obtenemos en el visor de resultados de **SPSS** las tablas de las Figuras 112 y 113.

	Casos					
	Válidos		Perdidos		Total	
	N	%	N	%	N	%
OPINIÓN1 * OPINIÓN2	30	100%	0	,0%	30	100%

Figura 112: Tabla-resumen de casos de las variables opinión1 y opinión2

En el cuadro de diálogo de la Figura 111 podemos seleccionar una variable de control en el tercer recuadro en blanco. Por cada categoría de esta tercera variable (la de control) se presenta una tabla de contingencia simple de las dos variables especificadas en Fila y Columna. Es posible especificar más de una variable de control. En el caso de que hubiese, por

		OPINIÓN2					Total
		1	2	3	4	5	
OPINIÓN1	1		2	3		1	6
	2	1		2	1	3	7
	3		2	3			5
	4	1	2	2	1	1	7
	5		1	1	2	1	5
Total		2	7	11	4	6	30

Figura 113: Tabla de contingencia de las variables opinión1 y opinión2

ejemplo, dos variables de control, se produciría una tabla de contingencia simple de las dos variables cuya relación se quiere examinar por cada combinación de categorías de las dos variables de control.

Por ejemplo, si en **Filas** seleccionamos la variable **opinión1**, en **Columnas** la variable **opinión2** y en el recuadro de las variables de control seleccionamos las variables **materia** y **vejez**, obtenemos en el visor de resultados las tablas de las Figuras 114 y 115.

MATERIA		OPINIÓN2					Total
		1	2	3	4	5	
E	OPINIÓN1 1			1			1
	2		1			1	2
	3			1			1
	4				1		1
	Total		1	1	2	1	3
G	OPINIÓN1 1			1			1
	2				1		1
	3				1		1
	5			1	1	2	4
	Total			2	3	2	7
H	OPINIÓN1 1				1		1
	3			1	1		2
	4		1				1
	Total		1	1	2		4
I	OPINIÓN1 1			1	1		2
	2				1		1
	5					1	1
	Total			1	2	3	6
L	OPINIÓN1 3				1		1
	4			1	1	1	3
	Total			1	2	1	4

Figura 114: Tabla de contingencia de opinión1 y opinión2 con la variable control materia

VEJEZ		OPINIÓN2					Total
		1	2	3	4	5	
1	OPINIÓN1 1		1	1			2
	2	1		1	1	2	5
	3		1	2			3
	4		1	1	1		3
	5		1	1			2
	Total		1	4	6	2	2
2	OPINIÓN1 1		1	1		1	3
	2			1		1	2
	4			1		1	2
	5				2	1	3
	Total		1	3	2	4	10
3	OPINIÓN1 3		1	1			2
	Total		1	1			2
4	OPINIÓN1 1			1			1
	4	1	1				2
Total		1	1	1			3

Figura 115: Tabla de contingencia de opinión1 y opinión2 con la variable control vejez

En el cuadro de diálogo de la Figura 111 hay dos opciones más, que son:

**Mostrar los gráficos de barras agrupados.** Se obtiene el gráfico de barras de la variable declarada en **Filas**, agrupado según la variable de **Columnas**.

**Suprimir tablas.** No se muestra ninguna tabla, sino sólo los estadísticos (en el caso de que por lo menos alguno de ellos haya sido solicitado).

En el cuadro de diálogo de la Figura 111 nos quedan tres botones por explicar: **Estadísticos**, **Casillas** y **Formato**.

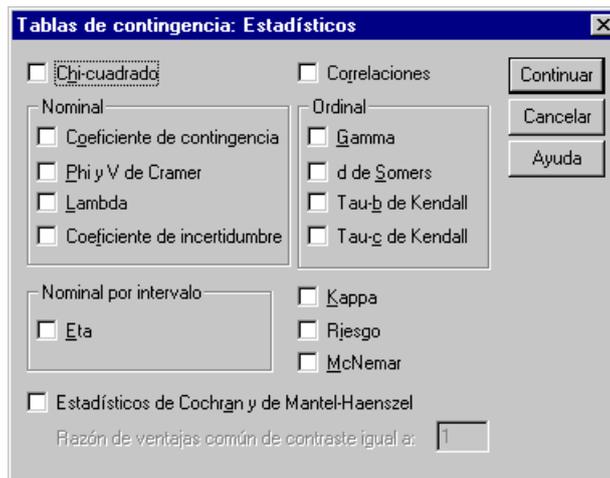


Figura 116: Cuadro de diálogo de la opción Estadísticos

Si pulsamos el botón **Estadísticos** nos aparece el cuadro de diálogo de la Figura 116. En este cuadro se pueden elegir los coeficientes de correlación que deseemos calcular. También con esta opción se realizan tests de hipótesis para contrastar la hipótesis de que el coeficiente de correlación en toda la población sea igual a cero. De entre todo lo que aparece, nos interesa marcar la opción **Chi-cuadrado**, que proporciona los resultados del contraste  $\chi^2$  de Pearson sobre independencia de las dos variables implicadas.

Por ejemplo, en **Filas** seleccionamos la variable **opinión1**, en **Columnas** la variable **opinión2** y al pulsar el botón **Estadísticos** seleccionamos **Chi-cuadrado**. Entonces obtenemos la tabla de contingencia de la Figura 113 y los resultados de la Figura 117.

	Valor	gl	Sig. asint. (bilateral)
Chi-cuadrado de Pearson	13,659 <sup>a</sup>	16	,624
Razón de verosimilitud	16,947	16	,389
Asociación lineal por lineal	,001	1	,971
N de casos válidos	30		

a. 25 casillas (100,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es ,33.

Figura 117: Contraste  $\chi^2$  de Pearson sobre independencia de las variables **opinión1** y **opinión2**

Comentaremos los resultados: En primer lugar, los contrastes *Razón de verosimilitud* y *Asociación lineal por lineal* no los hemos estudiado. En lo concerniente al contraste *Chi-cuadrado de Pearson* sobre independencia de las dos variables, el valor (muestral) del estadístico  $\chi^2$  es 13'659; los grados de libertad (*gl*) son 16, que es el resultado de la siguiente multiplicación: (número de filas-1)×(número de columnas-1); y el p-valor o nivel crítico (*Sig. asint. (bilateral)*) es 0'624 (mayor que  $\alpha = 0'05$ ), por lo que deberíamos aceptar la hipótesis nula de independencia de las variables **opinión1** y **opinión2**. Pero hay que tener en cuenta la observación que aparece en la parte inferior de la tabla, en la que se nos hace notar que el 100 % (más del 20 %) de las frecuencias esperadas ( $e_{ij}$ ) son inferiores a 5; por lo tanto, no podemos aplicar este contraste. Además, hay frecuencias esperadas menores que 1 (la menor es 0'33). En consecuencia, no podemos hacer caso de los resultados de este contraste para las variables seleccionadas, ya que no se cumplen las condiciones para poder aplicar dicho procedimiento.

Si pulsamos el botón **Casillas** del cuadro de diálogo de la Figura 111 nos aparece el nuevo cuadro de diálogo de la Figura 118. En este cuadro nos encontramos con tres bloques de opciones aditivas. En el bloque **Frecuencias** tenemos:

**Observadas.** Por defecto muestra las frecuencias absolutas observadas.

**Esperadas.** Son las frecuencias absolutas esperadas bajo la hipótesis de independencia entre ambas variables (útiles para realizar el test de independencia *chi-cuadrado*).

En el bloque **Porcentajes** tenemos:

**Fila.** Porcentaje de casos en cada casilla respecto de casos totales de la fila.

**Columna.** Análogo por columnas.

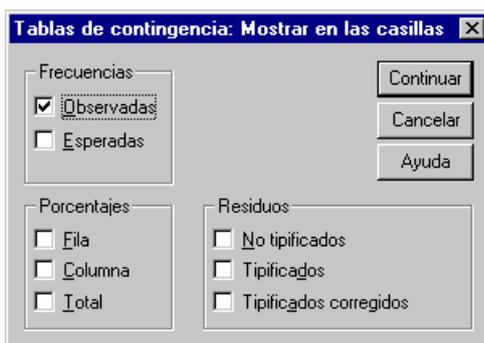


Figura 118: Cuadro de diálogo para especificar más cosas sobre las casillas de las tablas de contingencia

**Total.** Porcentaje de casos en cada casilla sobre el total de la tabla.

En el bloque **Residuos** sólo explicaré que la opción **No tipificados** presenta los resultados de las diferencias entre frecuencias absolutas observadas y las frecuencias absolutas esperadas.

Respecto del ejemplo que estamos comentando, si en **Frecuencias** seleccionamos **Observadas** y **Esperadas**; en **Porcentajes** seleccionamos **Total**; y en **Residuos** seleccionamos **No tipificados**; entonces obtenemos los resultados de la Figura 119.

		OPINIÓ2					Total	
		1	2	3	4	5		
OPINIÓ1	1	Recuento	0	2	3	0	1	6
		Frecuencia esperada	,4	1,4	2,2	,8	1,2	6,0
		% del total	,0%	6,7%	10%	,0%	3,3%	20,0%
		Residual	-,4	,6	,8	-,8	-,2	
	2	Recuento	1	0	2	1	3	7
		Frecuencia esperada	,5	1,6	2,6	,9	1,4	7,0
		% del total	3,3%	,0%	6,7%	3,3%	10%	23,3%
		Residual	,5	-1,6	-,6	,1	1,6	
	3	Recuento	0	2	3	0	0	5
		Frecuencia esperada	,3	1,2	1,8	,7	1,0	5,0
		% del total	,0%	6,7%	10%	,0%	,0%	16,7%
		Residual	-,3	,8	1,2	-,7	-1,0	
	4	Recuento	1	2	2	1	1	7
		Frecuencia esperada	,5	1,6	2,6	,9	1,4	7,0
		% del total	3,3%	6,7%	6,7%	3,3%	3,3%	23,3%
		Residual	,5	,4	-,6	,1	-,4	
5	Recuento	0	1	1	2	1	5	
	Frecuencia esperada	,3	1,2	1,8	,7	1,0	5,0	
	% del total	,0%	3,3%	3,3%	6,7%	3,3%	16,7%	
	Residual	-,3	-,2	-,8	1,3	,0		
Total	Recuento	2	7	11	4	6	30	
	Frecuencia esperada	2,0	7,0	11,0	4,0	6,0	30,0	
	% del total	6,7%	23%	37%	13%	20%	100%	

Figura 119: Nueva tabla de contingencia de las variables opinión1 y opinión2

Si pulsamos el botón **Formato** del cuadro de diálogo de la Figura 111 nos aparece el nuevo cuadro de diálogo de la Figura 120. En este cuadro sólo es posible reordenar las filas:

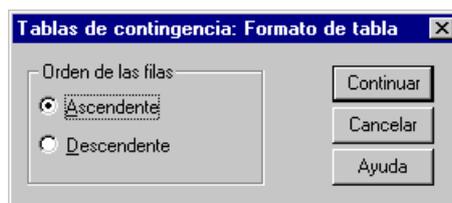


Figura 120: Cuadro de diálogo para especificar el formato de las tablas de contingencia

**Ascendente.** Se presentan los valores de la variable en orden ascendente (es la opción por defecto).

**Descendente.** Al contrario de la anterior.

### 5.3.1. Ejemplo de edición de una tabla de contingencia en el editor de datos

Supongamos que queremos saber si, en una determinada población, son independientes las variables “uso de la biblioteca” y “preferencia sobre distintos tipos de película”, según los datos de la siguiente muestra:

uso de la biblioteca	películas bélicas	películas de aventuras	películas melodramáticas
alto	16	8	2
medio	6	10	4
bajo	4	8	12



Figura 121: Datos del fichero Contingencia.sav

En primer lugar tenemos que introducir estos datos. Para ello, se crean dos variables, que se pueden denominar **uso** y **película**, de manera que entre ambas reflejen todas las posibles combinaciones entre sus modalidades. Los resultados de la variable **uso** serán: 1 (=“alto uso”), 2 (=“medio uso”) y 3 (=“bajo uso”). Los resultados de la variable **película** serán: 1 (=“bélicas”), 2 (=“aventuras”) y 3 (=“melodramáticas”). Además hay que crear otra variable, que llamaremos **frecuenc** con las frecuencias absolutas observadas en cada combinación de modalidades. Como se ha de reflejar todas las posibles combinaciones de los resultados de las dos variables, la matriz de datos ha de ser la de la Figura 121. No olvidemos grabar ahora este fichero de datos con el nombre **Contingencia.sav**.

			PELÍCULA			
			1	2	3	Total
USO	1	Recuento	16	8	2	26
		Frecuencia esperada	9,7	9,7	6,7	26,0
		Residual	6,3	-1,7	-4,7	
	2	Recuento	6	10	4	20
		Frecuencia esperada	7,4	7,4	5,1	20,0
		Residual	-1,4	2,6	-1,1	
	3	Recuento	4	8	12	24
		Frecuencia esperada	8,9	8,9	6,2	24,0
		Residual	-4,9	-9	5,8	
Total		Recuento	26	26	18	70
		Frecuencia esperada	26,0	26,0	18,0	70,0

Figura 122: Tabla de contingencia entre uso y película

A continuación, desde la ventana del *Editor de datos*, hay que seleccionar **Datos** ⇒ **Ponderar casos**. En el cuadro de diálogo resultante (Figura 123) hay que activar la opción **Ponderar casos mediante** y hay que seleccionar la variable **frecuenc** en el recuadro **Variable de frecuencia**.

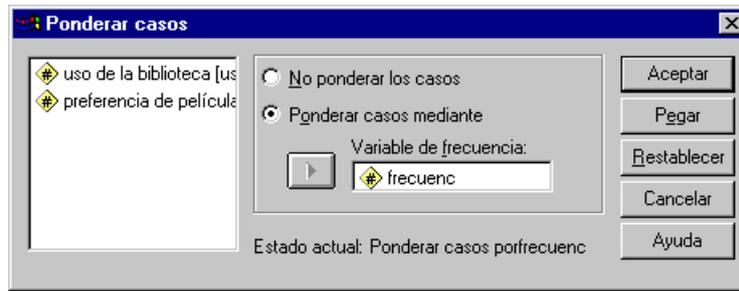


Figura 123: Cuadro de diálogo para ponderar casos

A partir de aquí se puede hacer todo lo anteriormente expuesto: Seleccionamos la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Tablas de contingencia**. En el cuadro de diálogo de la Figura 111 seleccionamos la variable **uso** en **Filas** y la variable **película** en **Columnas**. Pulsamos el botón **Estadísticos** y en el cuadro que aparece (Figura 116) activamos **Chi-cuadrado**. En el cuadro de diálogo de la Figura 111 podemos también pulsar el botón **Casillas** y activar (Figura 118) las opciones **Esperadas** (Frecuencias) y **No tipificados** (Residuos). Los resultados se muestran en las Figuras 122 y 124.

**Pruebas de chi-cuadrado**

	Valor	gl	Sig. asint. (bilateral)
Chi-cuadrado de Pearson	17,461 <sup>a</sup>	4	,002
Razón de verosimilitud	17,506	4	,002
Asociación lineal por lineal	15,199	1	,000
N de casos válidos	70		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 5,14.

Figura 124: Contraste chi-cuadrado sobre independencia entre las variables uso y película

Según los resultados (Figura 124), el nivel crítico (*Sig. asint. (bilateral)*) es igual a 0'002 (claramente menor que 0'05) por lo que rechazamos la hipótesis nula de que las variables **uso** y **película** son independientes. Por tanto, aceptamos que dichas variables son dependientes.

#### 5.4. Contraste de las rachas sobre aleatoriedad de la muestra

El contraste de las rachas sobre aleatoriedad de una muestra se realiza mediante **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **Rachas**. Esta prueba no puede utilizarse si los valores de la variable han sido ordenados en el archivo de datos. En **Contrastar variables**

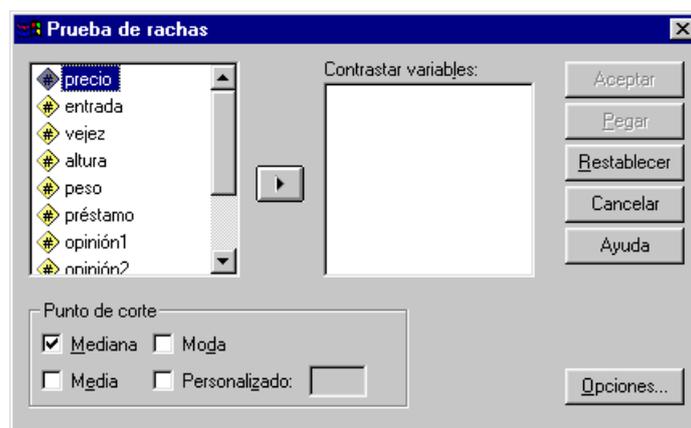


Figura 125: Cuadro de diálogo para el contraste de las rachas sobre aleatoriedad de la muestra

(Figura 125) se declaran las variables a analizar. Si hacemos *clic* en **Aceptar**, por defecto, las variables no dicotómicas se dicotomizan convenientemente a través de su mediana. De no ser tal criterio, el usuario puede establecer otros haciendo uso de las opciones de **Punto de corte**, de forma que en la primera categoría se agrupan los valores menores que el de corte especificado.

Aplicamos lo expuesto a la variable **precio** dicotomizada en función de la mediana. La conclusión (Figura 126) es que podemos aceptar la hipótesis nula de aleatoriedad de los resultados muestrales de la variable **precio**, ya que el nivel crítico (*Sig. asintót. (bilateral)*) es igual a 0'853 (claramente mayor que 0'05).

**Prueba de rachas**

	PRECIO
Valor de prueba <sup>a</sup>	7200,00
Casos < Valor de prueba	15
Casos ≥ Valor de prueba	15
Casos en total	30
Número de rachas	15
Z	-,186
Sig. asintót. (bilateral)	,853

a. Mediana

Figura 126: Contraste de las rachas sobre aleatoriedad de los resultados muestrales de la variable precio

**5.5. Contraste de Kolmogorov–Smirnov sobre bondad de ajuste. Normalidad**

Para hacer este contraste se selecciona **Analizar ⇒ Pruebas no paramétricas ⇒ K-S de 1 muestra** (Figura 127). Como siempre,

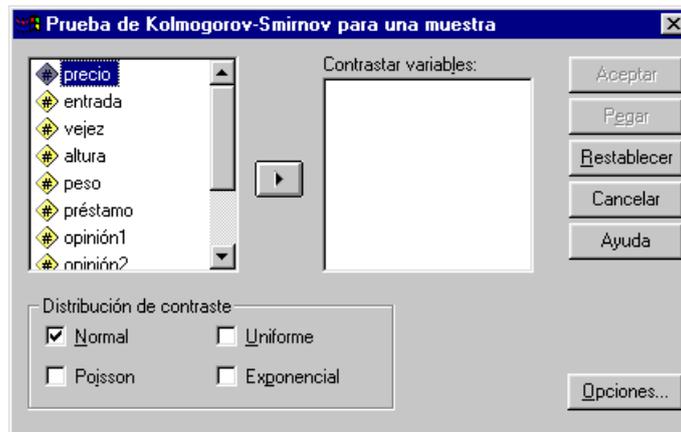


Figura 127: Cuadro de diálogo para el contraste de Kolmogorov–Smirnov sobre bondad de ajuste

en **Contrastar variables** son declaradas las variables que individualmente serán sometidas a análisis. Por defecto, haciendo clic en **Aceptar**, la distribución con la que se compara la observada es la Normal definida por la misma media y desviación típica que la empírica. En el caso de que no sea tal la distribución que el usuario desea tomar como segundo miembro de la hipótesis nula, se ofrecen otras tres opciones: **Uniforme**, **Poisson** y **Exponencial**.

Veamos ahora si se puede aceptar que las variables **precio**, **vejez**, **altura**, **peso** y **préstamo** siguen una distribución Normal. A la vista de los resultados (Figura 128) se puede aceptar que son normales las variables **precio**, **altura**, **peso** y **préstamo** y no lo es la variable **vejez** pues el nivel crítico (*Sig. asintót. (bilateral)*) es 0'015 (menor que 0'05).

**Prueba de Kolmogorov-Smirnov para una muestra**

		PRECIO	VEJEZ	ALTURA	PESO	PRÉSTAMO
N		30	30	30	30	30
Parámetros normales <sup>a,b</sup>	Media	8036,00	1,77	23.820	580,17	13,60
	Desviación típica	6015,43	,97	5.053	184,93	9,74
Diferencias más extremas	Absoluta	,130	,285	,212	,128	,169
	Positiva	,130	,285	,212	,128	,169
	Negativa	-,117	-,215	-,180	-,127	-,098
Z de Kolmogorov-Smirnov		,712	1,561	1,159	,699	,927
Sig. asintót. (bilateral)		,691	,015	,136	,714	,356

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Figura 128: Ejemplo del contraste de Kolmogorov–Smirnov sobre bondad de ajuste

**5.6. Contraste de homogeneidad con 2 muestras independientes**

El cuadro de diálogo inicial se logra con **Analizar ⇒ Pruebas no paramétricas ⇒ 2 muestras independientes** (Figura 129). En **Contrastar variables** se especifica la variable (o variables) objeto de estudio. Si se especifica más de una variable, éstas serán

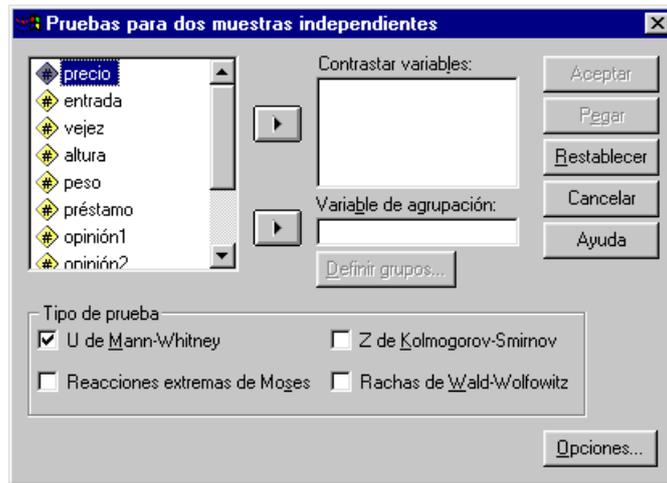


Figura 129: Cuadro de diálogo para el contraste de homogeneidad con dos muestras independientes

analizadas individualmente. En el recuadro **Variable de agrupación** se ha de especificar una variable dicotómica pues queremos comparar las distribuciones de la variable elegida en **Contrastar variables** para los dos grupos definidos por las dos categorías de la variable especificada en **Variable de agrupación**. A continuación se ha de pulsar el botón **Definir grupos** para especificar los dos grupos que queremos comparar (ver Figura 130). En el grupo **Tipo de prueba** se presentan, además del contraste **U de**



Figura 130: Cuadro de diálogo para definir los grupos en un contraste de homogeneidad con dos muestras independientes

**Mann-Whitney** ya conocido, otros como: **Reacciones extremas de Moises**, **Z de Kolmogorov-Smirnov** y **Rachas de Wald-Wolfowitz**.

Supongamos que queremos saber si la distribución de la variable **préstamo** es la misma en el grupo en que la variable **vejez** vale 1 (entre 1 y 3 años de antigüedad del libro), que en el grupo en que la variable **vejez** vale 2 (entre 4 y 6 años de antigüedad del libro). Para ello, en **Contrastar variables** colocamos la variable **préstamo** y en el recuadro **Variable de agrupación** elegimos la variable **vejez**. Pulsamos **Definir grupos** y, en el cuadro que aparece, escribimos 1 en **Grupo 1** y 2 en **Grupo 2**. Pulsamos **Continuar** y en el cuadro inicial sólo dejamos seleccionada la opción **U de Mann-Whitney**. Tras pulsar **Aceptar** obtenemos los resultados de la Figura 131. Como el nivel crítico (*Sig. asintót. (bilateral)*) es igual a 0'039 (menor que 0'05) rechazamos la hipótesis nula de homogeneidad; es decir, el número de préstamos no es el mismo en el grupo de libros cuya antigüedad está comprendida entre 1 y 3 años, que en el grupo de libros cuya antigüedad está comprendida entre 4 y 6 años.

#### Rangos

	VEJEZ	N	Rango promedio	Suma de rangos
PRÉSTAMO	1	15	10,53	158,00
	2	10	16,70	167,00
	Total	25		

#### Estadísticos de contraste<sup>b</sup>

	PRÉSTAMO
U de Mann-Whitney	38,000
W de Wilcoxon	158,000
Z	-2,060
Sig. asintót. (bilateral)	,039
Sig. exacta [2*(Sig. unilateral)]	,041 <sup>a</sup>

a. No corregidos para los empates.

b. Variable de agrupación: VEJEZ

Figura 131: Ejemplo de un contraste de homogeneidad con dos muestras independientes

Si tuviésemos que introducir, en el editor de datos de *SPSS*, los resultados de dos muestras independientes para comprobar si las dos poblaciones son homogéneas, habría que hacerlo de forma totalmente análoga a la manera descrita en el apartado 4.2.1: la primera columna va a contener todos los resultados de las dos muestras, una debajo de la otra; y la segunda columna de datos va a tener un 1 para los resultados que provengan de la muestra primera, o un 2 para los resultados que provengan de la muestra segunda.

### 5.7. Contraste de homogeneidad con más de dos muestras independientes

El cuadro de diálogo inicial se logra con **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **K muestras independientes** (Figura 132). En **Contrastar variables** se especifica la variable (o variables) objeto de estudio. Si se especifica más de una variable, éstas



Figura 132: Cuadro de diálogo para el contraste de homogeneidad con más de dos muestras independientes

serán analizadas individualmente. Queremos comparar las distribuciones de la variable elegida en **Contrastar variables** para los distintos grupos definidos por las categorías de la variable especificada en **Variable de agrupación**. A continuación se ha de pulsar el botón **Definir grupos** para especificar los grupos que queremos comparar (ver Figura 133). En el grupo **Tipo de prueba** se

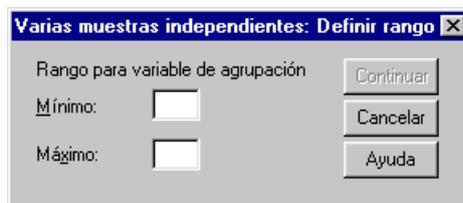


Figura 133: Cuadro de diálogo para definir los grupos en un contraste de homogeneidad con más de dos muestras independientes

presenta, además del contraste **H de Kruskal–Wallis**, el test de la **Mediana** (que prueba si las *k* muestras independientes proceden de poblaciones con la misma mediana).

Supongamos que queremos saber si la distribución de la variable **préstamo** es la misma en los cuatro grupos en que tenemos dividida la variable **vejez** (1=entre 1 y 3 años, 2=entre 4 y 6 años, 3=entre 7 y 9 años, 4=10 años o más). Para ello, en **Contrastar variables** colocamos la variable **préstamo** y en el recuadro **Variable de agrupación** elegimos la variable **vejez**. Pulsamos **Definir grupos** y, en el cuadro que aparece, escribimos 1 en **Mínimo** y 4 en **Máximo**. Pulsamos **Continuar** y en el cuadro inicial dejamos seleccionada la opción **H de Kruskal–Wallis**. Tras pulsar **Aceptar** obtenemos los resultados de la Figura 134. Como el nivel crítico (*Sig. asintót.*) es igual a 0'033 (menor que 0'05) rechazamos la hipótesis nula de homogeneidad; es decir, el número de préstamos no es el mismo en los cuatro grupos; es decir, el número de veces que un libro es prestado depende de la antigüedad de dicho libro.

Si tuviésemos que introducir, en el editor de datos de *SPSS*, los resultados de *k* muestras independientes para comprobar si las *k* poblaciones son homogéneas, habría que hacerlo de forma totalmente análoga a la manera descrita en el apartado 4.2.1: los resultados de las *k* muestras se ponen todos en la primera columna (una muestra debajo de otra) pero en esta ocasión la variable **grupo** tendría por resultados: 1, 2, 3, . . . , *k*, dependiendo de a qué muestra pertenece cada resultado de los que están en la columna primera.

### 5.8. Contraste de homogeneidad con 2 muestras relacionadas

El cuadro de diálogo inicial se logra con **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **2 muestras relacionadas** (Figura 135).

En primer lugar debemos seleccionar (marcar) un par de variables de la lista; veremos como éstas quedarán escritas automáticamente junto a **Variable 1** y **Variable 2**. Después pulsamos el botón **▶**. En **Tipo de prueba** se presentan, además del

**Rangos**

	VEJEZ	N	Rango promedio
PRÉSTAMO	1	15	13,27
	2	10	20,40
	3	2	22,00
	4	3	6,00
	Total	30	

**Estadísticos de contraste<sup>a,b</sup>**

	PRÉSTAMO
Chi-cuadrado	8,707
gl	3
Sig. asintót.	,033

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: VEJEZ

Figura 134: Ejemplo de un contraste de homogeneidad con más de dos muestras independientes

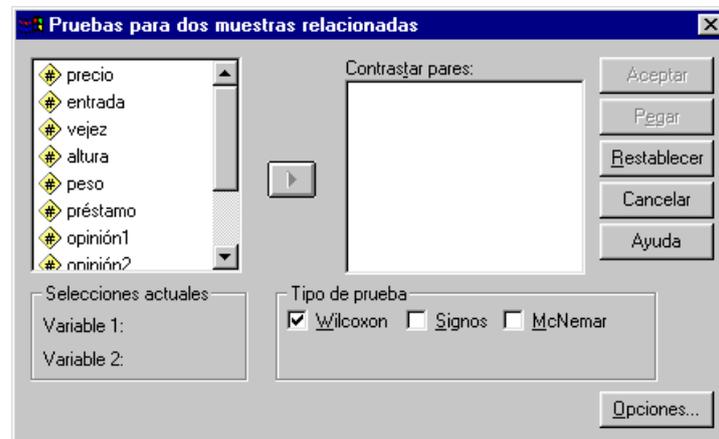


Figura 135: Cuadro de diálogo para el contraste de homogeneidad con dos muestras dependientes

conocido contraste de homogeneidad de **Wilcoxon**, el contraste de los **Signos** y el de **McNemar** (que sólo es válido cuando ambas variables son dicotómicas).

**Rangos**

		N	Rango promedio	Suma de rangos
OPINIÓN3 - OPINIÓN2	Rangos negativos	10 <sup>a</sup>	11,75	117,50
	Rangos positivos	15 <sup>b</sup>	13,83	207,50
	Empates	5 <sup>c</sup>		
	Total	30		

a. OPINIÓN3 < OPINIÓN2

b. OPINIÓN3 > OPINIÓN2

c. OPINIÓN2 = OPINIÓN3

**Estadísticos de contraste<sup>b</sup>**

	OPINIÓN3 - OPINIÓN2
Z	-1,230 <sup>a</sup>
Sig. asintót. (bilateral)	,219

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Figura 136: Ejemplo del contraste de Wilcoxon sobre homogeneidad con dos muestras dependientes

A partir de la muestra de datos contenida en el fichero **Libros.sav**, veamos si la opinión de los expertos 2 y 3 respecto de la encuadernación de los libros se puede considerar la misma en la población total de libros. Notemos que, efectivamente, las dos muestras son dependientes, pues realmente son la misma muestra (a la cual se le ha observado dos variables distintas). Para ello, marcamos las variables **opinión2** y **opinión3**, pulsamos el botón  y dejamos seleccionada la opción **Wilcoxon**. Tras pulsar **Aceptar** obtenemos los resultados de la Figura 136. Como el nivel crítico es 0'219 (mayor que 0'05) entonces se puede aceptar la hipótesis de homogeneidad; es decir, que no hay diferencia de criterio entre los expertos 2 y 3 respecto de la encuadernación de los libros.



Figura 137: Cuadro de diálogo para el contraste de homogeneidad con más de dos muestras dependientes

**Rangos**

	Rango promedio
OPINIÓN1	1,78
OPINIÓN2	1,95
OPINIÓN3	2,27

**Estadísticos de contraste<sup>a</sup>**

N	30
Chi-cuadrado	4,173
gl	2
Sig. asintót.	,124

a. Prueba de Friedman

Figura 138: Ejemplo del contraste de Friedman sobre homogeneidad con más de dos muestras dependientes

## 5.9. Contraste de homogeneidad con más de dos muestras relacionadas

El cuadro de diálogo inicial se logra con **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **K muestras relacionadas** (Figura 137).

Para declarar las variables que serán objeto de análisis debemos seleccionarlas (marcarlas) en la lista que aparece en el cuadro de diálogo. Después pulsamos el botón . En **Tipo de prueba** se presentan, además del conocido contraste de homogeneidad de **Friedman**, el contraste **W de Kendall** y el **Q de Cochran** (que sólo es válido cuando todas las variables seleccionadas son dicotómicas).

A partir de la muestra de datos contenida en el fichero **Libros.sav**, veamos si la opinión de los tres expertos en la encuadernación de los libros se puede considerar la misma en la población. Marcamos las variables **opinión1**, **opinión2** y **opinión3**, pulsamos el botón  y dejamos seleccionada la opción **Friedman**. Los resultados aparecen en la Figura 138. Debido a que el nivel crítico es 0'124 (mayor que 0'05) entonces se puede aceptar la hipótesis de homogeneidad; es decir, que no hay diferencia de criterio entre los tres expertos respecto de la encuadernación de los libros.