

**ESTADÍSTICA**  
**Licenciatura en Documentación**  
**Curso Académico 2007/2008**  
**Manual de prácticas de ordenador con SPSS 15**  
**Departamento de Estadística e Investigación Operativa**  
**Autora: Dra. Josefa Marín Fernández**

## Índice de contenidos

<b>1. Normas generales de SPSS</b>	<b>4</b>
1.1. Introducción	4
1.2. Barra de menús	4
1.3. Presentación de nuestro ejemplo	5
1.4. Definición de la primera variable de nuestro ejemplo	5
1.5. Grabación de archivos de datos	9
1.6. Definición del resto de las variables de nuestro ejemplo	9
1.7. Introducción de los datos a través del teclado	12
1.8. Lectura de archivos de datos propios de SPSS	12
1.9. Edición y modificación de los datos	12
1.10. Modificación de la apariencia del editor de datos	15
1.11. Impresión de los datos	16
1.12. Importación de archivos de datos	16
1.13. Transformaciones en el archivo de datos	16
1.13.1. Transformación de variables sin utilizar el condicional	16
1.13.2. Transformación condicional de variables	19
1.13.3. Clasificación de variables mediante la opción <i>Transformar</i> ⇒ <i>Recodificar ...</i>	20
1.13.4. Clasificación de variables mediante la opción <i>Transformar</i> ⇒ <i>Agrupación visual</i>	22
1.13.5. Ordenación de los resultados de una variable	26
1.13.6. Ponderación de los casos	28
<b>2. Estadística descriptiva unidimensional</b>	<b>30</b>
2.1. Distribución de frecuencias	30
2.2. Observaciones sobre la presentación, modificación y grabación de los resultados en la ventana del visor SPSS	30
2.2.1. Cómo guardar y recuperar una ventana del visor SPSS	30
2.2.2. Cómo controlar la visualización de la información sobre variables y valores de datos en las tablas pivote y en los titulares	31
2.2.3. Cómo establecer el aspecto de las tablas pivote por defecto	31
2.2.4. Cómo cambiar las especificaciones del visor SPSS por defecto	32
2.3. Medidas descriptivas de los datos	33
2.3.1. Mediante la opción <i>Analizar</i> ⇒ <i>Estadísticos descriptivos</i> ⇒ <i>Frecuencias</i>	33
2.3.2. Mediante la opción <i>Analizar</i> ⇒ <i>Estadísticos descriptivos</i> ⇒ <i>Descriptivos</i>	35
2.3.3. Mediante la opción <i>Analizar</i> ⇒ <i>Estadísticos descriptivos</i> ⇒ <i>Explorar</i>	37
2.4. Representaciones gráficas unidimensionales	37
2.4.1. Representaciones gráficas mediante la opción <i>Analizar</i> ⇒ <i>Estadísticos descriptivos</i> ⇒ <i>Frecuencias</i>	37
2.4.2. Cómo cambiar las especificaciones de las representaciones gráficas por defecto	39
2.4.3. Creación de diagramas de barras mediante el menú de <i>Gráficos</i>	40
2.4.4. Cuadro de diálogo de <i>Cambiar estadístico</i>	46
2.4.5. Cuadro de diálogo de <i>Opciones</i>	47
2.4.6. Cuadro de diálogo de <i>Títulos</i>	48
2.4.7. Creación de histogramas mediante el menú de <i>Gráficos</i>	48
2.4.8. Modificación de gráficos	49
<b>3. Relación entre dos variables</b>	<b>53</b>
3.1. Relación entre variables cuantitativas. Diagrama de dispersión	53
3.2. Modificación del diagrama de dispersión	53
3.3. Correlación simple	56
3.4. Regresión lineal	59
<b>4. Probabilidad. Variables aleatorias</b>	<b>61</b>
4.1. Función de probabilidad y función de densidad	61
4.2. Función de distribución (probabilidad acumulada)	62
4.3. Inversa de la función de distribución (percentiles o cuantiles)	64

<b>5. Contrastes paramétricos</b>	<b>66</b>
5.1. Contraste para una media. Intervalo de confianza para la media . . . . .	66
5.2. Comparación de dos medias con muestras independientes . . . . .	67
5.2.1. Ejemplo de edición de unos datos para hacer la comparación de dos medias con muestras independientes . . . . .	69
5.3. Comparación de dos medias con muestras relacionadas . . . . .	70
<b>6. Contrastes no paramétricos</b>	<b>73</b>
6.1. Contraste sobre una proporción . . . . .	73
6.2. Contraste $\chi^2$ de Pearson sobre proporciones . . . . .	73
6.3. Contraste $\chi^2$ sobre independencia de dos variables . . . . .	75
6.3.1. Ejemplo de edición de una tabla de contingencia en el editor de datos . . . . .	79
6.4. Contraste de las rachas sobre aleatoriedad de la muestra . . . . .	80
6.5. Contraste de Kolmogorov–Smirnov sobre bondad de ajuste. Normalidad . . . . .	80
6.6. Contraste de homogeneidad con 2 muestras independientes . . . . .	81
6.7. Contraste de homogeneidad con más de dos muestras independientes . . . . .	83
6.8. Contraste de homogeneidad con 2 muestras relacionadas . . . . .	83
6.9. Contraste de homogeneidad con más de dos muestras relacionadas . . . . .	85

# 1. Normas generales de SPSS

## 1.1. Introducción

Las prácticas se van a realizar con el programa *SPSS para Windows*, versión 15 (en español). Para ejecutar el programa debemos seleccionar **Inicio** ⇒ **Programas** ⇒ **SPSS 15 para Windows**. Lo primero que aparece es la ventana básica de esta aplicación, o para ser más precisos, la *ventana del editor de datos* (Figura 1). Como en cualquier otra aplicación Windows,

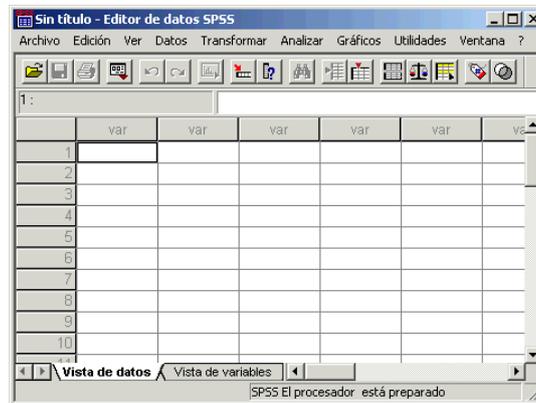


Figura 1: Ventana del editor de datos

la *ventana del editor de datos* puede modificarse en cuanto al tamaño y a la disposición de sus elementos. 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 menús** 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.
- 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 dos “carpetas”. La que habitualmente aparece encima es la carpeta denominada **Vista de datos** y la que habitualmente aparece debajo es la carpeta denominada **Vista de variables**. La carpeta **Vista de datos** muestra los valores de datos reales o las etiquetas de valor definidas. La carpeta **Vista de variables** muestra la información de definición de variable, que incluye las etiquetas de la variable definida y de valor, tipo de dato (por ejemplo, de cadena, fecha y numérico), escala de medida (nominal, ordinal o de escala) y los valores perdidos definidos por el usuario.
- 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 *El procesador está 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 archivos 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 insertar variables, así como efectuar modificaciones en los archivos de datos: seleccionar, añadir, ponderar, etc.

**Transformar:** Aquí se encuentran todas las opciones relativas a la modificación y generación de nuevas variables.

**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 archivos 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 se selecciona la opción **Archivo** ⇒ **Salir** o se pulsa el botón de la esquina superior derecha de la ventana (botón ).

### 1.3. 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 Figura 2, que se refieren al estudio de diversas variables en una muestra de libros de una biblioteca universitaria. **De momento no introduciremos ningún dato; sólo leeremos lo concerniente al archivo de datos que posteriormente crearemos.** Las variables son: materia (disciplina científica a la que pertenece el libro); precio (en euros); entrada (fecha de entrada del libro en la biblioteca); 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).

### 1.4. Definición de la primera variable de nuestro ejemplo

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:

- 1) Seleccionamos la carpeta **Vista de variables** haciendo *clic* en su correspondiente pestaña (parte inferior del *editor de datos*). Nos aparece, entonces, la ventana de la Figura 3.
- 2) En la casilla intersección de la fila correspondiente a nuestra variable (fila 1) con la columna correspondiente a **Nombre** escribimos el nombre de nuestra primera variable, que es, como ya hemos dicho, **materia**

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 debe evitar acabar los nombres de variable con subrayado (para evitar conflictos con las variables creadas automáticamente por algunos procedimientos).
- La longitud del nombre no debe exceder los 64 bytes. Normalmente, 64 bytes suelen equivaler a 64 caracteres en idiomas de un solo byte (por ejemplo, inglés, francés, alemán, español, italiano, hebreo, ruso, griego, árabe, tailandés) y a 32 caracteres en los idiomas de dos bytes (por ejemplo, japonés, chino, coreano).
- No se pueden utilizar espacios en blanco ni caracteres especiales (como, por ejemplo, !, ?, ' o \*).
- Las palabras reservadas no se pueden utilizar como nombres de variable. Las palabras reservadas son: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.
- Los nombres de variable se pueden definir combinando de cualquier manera caracteres en mayúsculas y en minúsculas, esta distinción entre mayúsculas y minúsculas se conserva en lo que se refiere a la presentación. Cuando es necesario dividir los nombres largos de variable en varias líneas en los resultados, *SPSS* intenta dividir las líneas aprovechando los subrayados, los puntos y los cambios de minúsculas a mayúsculas.

	materia	precio	entrada	altura	peso	préstamo	opinión1	opinión2	opinión3
1	E	110	03.05.1999	20.5	325	32	2	1	1
2	G	30	06.12.2000	25.6	890	18	5	4	5
3	E	75	25.10.2000	20.0	415	9	4	5	5
4	I	45	03.09.2001	20.0	400	15	2	3	4
5	L	32	12.03.2000	18.8	515	2	3	3	3
6	I	69	26.06.2002	25.0	650	36	1	5	3
7	H	30	15.10.1999	30.6	790	5	3	3	1
8	H	34	26.11.2001	28.9	890	9	1	3	2
9	L	42	07.07.2000	20.1	320	7	4	2	1
10	I	46	04.05.2002	20.0	420	12	2	5	3
11	E	53	06.10.2000	23.5	620	42	3	2	2
12	E	97	12.12.2001	23.5	720	20	4	3	4
13	G	57	12.05.2002	30.9	850	10	3	3	5
14	I	51	13.11.1999	20.5	620	9	1	3	2
15	E	43	15.02.2002	22.0	475	15	2	5	5
16	L	36	01.02.2001	20.0	320	1	4	4	3
17	L	42	11.11.1999	21.0	380	15	4	3	5
18	G	37	07.06.2000	35.9	790	16	5	4	4
19	G	82	23.04.1998	33.0	780	9	5	3	5
20	I	48	21.10.2002	20.0	500	16	1	2	4
21	E	45	03.03.2002	25.9	550	11	2	5	4
22	G	58	11.08.1998	25.0	820	8	2	3	3
23	H	66	04.04.1999	19.8	360	4	3	2	4
24	H	40	23.09.2001	25.3	490	4	4	1	4
25	*	35	15.12.2000	22.8	480	2	4	2	4
26	E	45	06.09.1999	19.8	390	23	1	3	5
27	E	105	12.03.2002	20.0	520	16	2	4	3
28	G	61	08.05.2000	20.0	650	13	1	2	3
29	I	57	04.10.2001	20.0	690	20	5	5	4
30	G	51	06.09.1998	36.2	785	9	5	2	5

Figura 2: Datos que, posteriormente, vamos a introducir y grabar con el nombre **Libros.sav**

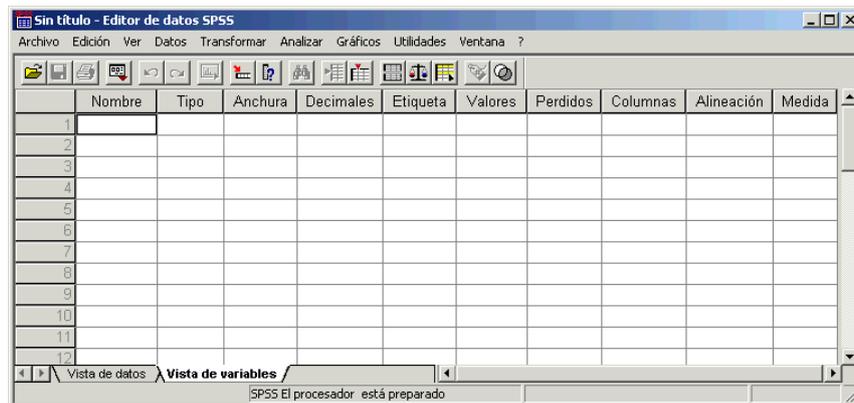


Figura 3: Vista de variables

- 3) En la carpeta **Vista de Variables** (Figura 3) hacemos *clic* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable (fila 1) con la columna correspondiente a **Tipo**. Aparece, entonces, el cuadro de diálogo de la Figura 4. Aquí se especifica el tipo 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 de este cuadro de diálogo 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 variables disponibles son:

**Numérica:** En este formato, la separación decimal se hace mediante una coma (abajo). En la anchura de los datos cuenta el signo (+ o -), si lo ponemos, y cuenta la coma de la separación decimal.



Figura 4: Cuadro de dialogo para definir el tipo de variable

**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:** Los valores de la variable se muestran con una E (o una D) intercalada y un exponente con signo que representa una potencia de base diez. El *editor de datos* acepta para estas variables valores numéricos con o sin el exponente. El exponente puede aparecer precedido por una E o una D con un signo opcional, o bien sólo por el signo. En este formato 347E-5 significa 347 multiplicado por 10 elevado a -5; es decir,  $347 \times 10^{-5} = 347 \times \frac{1}{100000} = 0'00347$ . Por ejemplo, es lo mismo 123; 1,23E2; 1,23D2; 1,23E+2 e incluso 1,23+2.

**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**  $\Rightarrow$  *Carpeta 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 (definición de la variable **materia**), activamos **Cadena** del cuadro de diálogo (Figura 4), 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 **Aceptar** para volver a la **Vista de Variables** (Figura 3).

- 4) Como el nombre que se le ha dado a la variable podría no ser suficientemente explícito, debemos poner una etiqueta explicativa a esta variable. Para ello, en la carpeta **Vista de variables** (Figura 3) hacemos *click* en la casilla intersección de la fila correspondiente a nuestra variable (fila 1) con la columna correspondiente a **Etiqueta** y escribimos *disciplina científica del libro*.
- 5) No sólo se puede poner una explicación o etiqueta al nombre de la variable sino también a cada uno de sus valores o resultados. 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).

Para poner etiquetas a los resultados de la variable **materia**, en la carpeta **Vista de Variables** (Figura 3) hacemos *click* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable (fila 1) con la columna correspondiente a **Valores**. Aparece, entonces, el cuadro de diálogo de la Figura 5.

- Junto a **Valor** escribimos *E*, junto a **Etiqueta** escribimos *estadística* y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos *G*, junto a **Etiqueta** escribimos *geografía* y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos *I*, junto a **Etiqueta** escribimos *informática* y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos *L*, junto a **Etiqueta** escribimos *literatura* y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos *H*, junto a **Etiqueta** escribimos *historia* y pulsamos el botón **Añadir**.

Una vez finalizado el proceso se pulsa **Aceptar**.

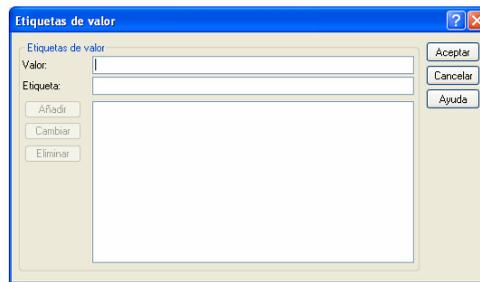


Figura 5: Cuadro de diálogo para definir etiquetas de valores

- 6) En la Figura 2 del apartado 1.3 podemos ver que para el libro número 25 no se sabe el resultado de la variable **materia**. Esto significa que dicho resultado es un valor perdido o ausente. En **SPSS** tenemos la posibilidad de 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. Nosotros vamos a representar los valores perdidos de nuestra variable **materia** por un asterisco (\*). Para ello, en la carpeta **Vista de Variables** (Figura 3) hacemos *click* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable (fila 1) con la columna correspondiente a **Perdidos** y nos aparece el cuadro de diálogo de la Figura 6. En dicho cuadro de diálogo debemos señalar la opción **Valores perdidos discretos** y dentro del primer campo

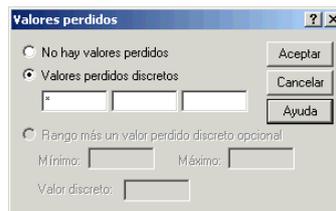


Figura 6: Cuadro de diálogo para definir los valores perdidos de una variable

escribir un asterisco (\*). A continuación pulsamos el botón **Aceptar**.

- 7) En la carpeta **Vista de Variables** (Figura 3) tenemos la opción de cambiar el formato de columna de la variable. Esto se refiere al número de dígitos necesarios para que en la matriz de datos se vean tanto los resultados de la variable como el nombre de dicha variable. En el caso de la variable **materia** sus resultados tienen un sólo dígito pero su nombre ocupa 7 dígitos. Por tanto, debemos poner un formato de columna igual o mayor que 7. Nosotros pondremos 7. Para ello, hacemos *click* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Columnas** y vemos que aparecen unas pequeñas flechas grises que nos sirven para aumentar o disminuir el número de dígitos. Por defecto, se nos ofrecen 8 dígitos. Nosotros bajaremos a 7, pulsando una vez la flecha que indica hacia abajo.

Con esta opción se controla el ancho de columnas en la carpeta **Vista de datos**. Los anchos de columna también se pueden cambiar en dicha carpeta pulsando y arrastrando los bordes de la columna. Los formatos de columna afectan sólo a la presentación de valores en la mencionada carpeta. Al cambiar el ancho de columna no se cambia el ancho de una variable, definido en **Tipo**. Si el ancho real de un valor es más ancho que la columna, aparecerán varios asteriscos (\*\*\*) en la carpeta **Vista de datos**; eso significa que los resultados no caben y, por tanto, tenemos que aumentar la cantidad de dígitos indicados en **Columnas** (en la carpeta **Vista de variables**).

- 8) La opción señalada como **Alineación** en la carpeta **Vista de Variables** (Figura 3) controla la presentación de los valores de los datos y/o de las etiquetas de valor en la **Vista de datos**. La alineación por defecto es derecha para las variables numéricas e izquierda para las variables de cadena. Esta configuración afecta sólo a la presentación en la **Vista de datos**.

Si queremos, por ejemplo, que los resultados de nuestra variable **materia** aparezcan alineados a la derecha, entonces en la carpeta **Vista de Variables** (Figura 3) hacemos *click* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Alineación**. Aparece, entonces, una pequeña flecha gris que, al pulsar, nos ofrece las tres opciones: *Izquierda*, *Derecha*, *Centrado*; de las cuales elegimos **Derecha**.

- 9) La opción señalada como **Medida** en la carpeta **Vista de Variables** (Figura 3) se refiere al tipo de variable: *Esca-la*=variable cuantitativa (discreta o continua), *Ordinal*=cualitativa ordinal o *Nominal*=cualitativa nominal.

Para señalar que nuestra variable **materia** es cualitativa nominal, en la carpeta **Vista de Variables** (Figura 3) hacemos *clic* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Medida**. Aparece, entonces, una pequeña flecha gris que, al pulsar, nos ofrece las tres opciones mencionadas, de las cuales elegimos **Nominal**.

## 1.5. Grabación de archivos de datos

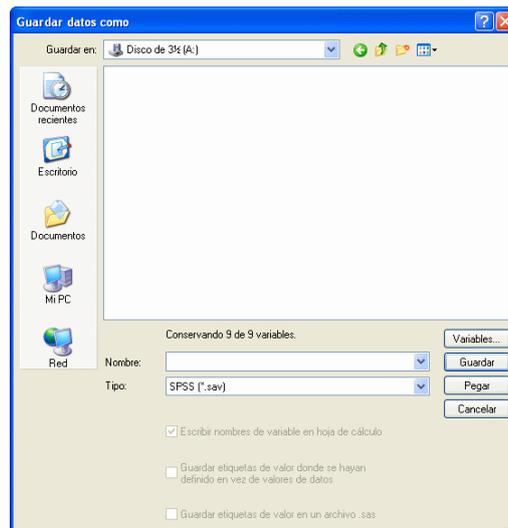


Figura 7: Cuadro de diálogo para grabar un archivo de datos

Aunque sólo hemos definido una variable y ni siquiera hemos introducido sus resultados, si no queremos perder el trabajo realizado hasta ahora (debido a un corte de electricidad o cualquier otra causa), podemos grabar el que será nuestro archivo 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 archivos de datos: **Guardar** y **Guardar como**.

Al igual que en otras muchas aplicaciones *Windows*, la opción **Guardar** sirve para grabar un archivo 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 archivo 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 archivo 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 archivo o al ir a modificar algunos de sus parámetros de grabación. Al activarla aparece un cuadro de diálogo como el de la Figura 7. En el mismo debemos especificar, mediante los procedimientos habituales, el nombre del archivo, la unidad y la carpeta de la misma en la que queremos que se grabe.

Con respecto a la posibilidad de grabar el archivo con un tipo distinto al de *SPSS para Windows* (.sav), esto lo veremos posteriormente. En consecuencia, en **Tipo** del cuadro de diálogo de la Figura 7 dejamos inalterada la opción seleccionada por defecto que es la de un archivo de datos de *SPSS para Windows* (.sav). En este cuadro de diálogo 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 (en la unidad A, en una memoria extraíble o en los contenidos de alumnos de SUMA) nuestro archivo de datos. Recordemos que el nombre de este archivo es **Libros.sav**

## 1.6. Definición del resto de las variables de nuestro ejemplo

El procedimiento descrito en el apartado 1.4 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 insertar una**

variable entre otras ya existentes.

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

- ◇ En primer lugar, seleccionamos la carpeta **Vista de variables** haciendo *clic* en su correspondiente pestaña (parte inferior del *editor de datos*) (Figura 3).
- ◇ En la casilla intersección de la fila correspondiente a nuestra variable (fila 2) con la columna correspondiente a **Nombre** escribimos **entrada**
- ◇ Hacemos *clic* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable (fila 2) con la columna correspondiente a **Tipo**. En el cuadro de diálogo que aparece seleccionamos **Fecha** y de las opciones que se muestran seleccionamos, por ejemplo, el formato **dd.mm.yyyy**, que es el quinto que se ofrece (ver Figura 8). Luego pulsamos **Aceptar**.



Figura 8: Cuadro de dialogo para definir el tipo fecha

- ◇ Como el nombre que se le ha dado a la variable podría no ser suficientemente explícito, deberíamos poner una etiqueta explicativa a esta segunda variable. Para ello, en la carpeta **Vista de variables**, concretamente en la casilla intersección de la fila correspondiente a nuestra variable (fila 2) con la columna correspondiente a **Etiqueta**, escribimos *fecha de entrada del libro en la biblioteca*.
- ◇ No es preciso especificar el formato de columna en **Columnas** de la carpeta **Vista de variables** dado que al seleccionar el formato de fecha **dd.mm.yyyy** el ancho de la columna implícito será de 10 (los puntos también cuentan).
- ◇ Así, la variable **entrada** ha sido completamente definida. Ahora se puede volver a grabar el que será nuestro archivo de datos denominado **Libros.sav** (*Archivo* ⇒ *Guardar*).

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

- ♡ En la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Nombre** (de la carpeta **Vista de variables**) escribimos **altura**
- ♡ Hacemos *clic* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Tipo**.

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.

Debido a que la introducción de los resultados (apartado 1.7) la haremos mediante el teclado numérico, y en dicho teclado no está la separación decimal mediante la coma, pero sí está la separación decimal mediante el punto, entonces elegiremos la opción **Coma**. En **Anchura** debemos escribir 4 (pues son dos cifras para la parte entera, más una coma, más un decimal) y en **Cifras decimales** escribimos 1

- ♡ Ahora pondremos una etiqueta explicativa a esta tercera variable. Para ello, en la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Etiqueta**, escribimos *altura del libro, en centímetros*.
- ♡ Los resultados de la variable **altura** tienen 4 dígitos y su nombre ocupa 6 dígitos. Por tanto, vamos a poner un formato de columna con una anchura igual a 6. Para ello, hacemos *clic* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Columnas** y seleccionamos un 6.
- ♡ Volvemos a grabar los datos en **Libros.sav** (*Archivo* ⇒ *Guardar*).

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

- ♣ En **Nombre** escribimos **peso**
  - ♣ En **Tipo** seleccionamos **Numérica**. Como los resultados de la variable **peso** tienen, como máximo, tres dígitos, sin decimales, entonces en **Anchura** escribimos un 3 y en **Cifras decimales** escribimos un 0.
  - ♣ En **Etiqueta** escribimos *peso del libro, en gramos*.
  - ♣ En **Columnas** seleccionamos 4 (pues así se verá el nombre completo de la variable y sus resultados).
  - ♣ Volvemos a grabar los datos en **Libros.sav**
- 4) 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 **Nombre** escribimos **préstamo**
  - En **Tipo** seleccionamos **Numérica**. Como los resultados de la variable **préstamo** tienen, como máximo, dos dígitos, sin decimales, entonces en **Anchura** escribimos un 2 y en **Cifras decimales** escribimos un 0.
  - En **Etiqueta** escribimos *número de veces que se ha prestado el libro en el último año*.
  - En **Columnas** seleccionamos 8 (pues así se verá el nombre de la variable y sus resultados).
  - Volvemos a grabar los datos en **Libros.sav**
- 5) 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 definiremos la primera de las tres variables y luego copiaremos y pegaremos todos los atributos.
- Para definir las tres últimas variables de nuestro ejemplo, seguiremos los siguientes pasos:
- a) Definimos la primera de las tres variables:
    - En **Nombre** escribimos **opinión1**
    - En **Tipo** seleccionamos **Numérica** con una **Anchura** igual a 1 y un número de **Cifras decimales** igual a 0.
    - En **Etiqueta** escribimos *opinión del primer experto respecto de la encuadernación del libro*.
    - En **Valores** asignamos las siguientes etiquetas de resultados (ver la Figura 5): 1=muy mala, 2=mala, 3=regular, 4=buena, 5=muy buena.
    - En **Columnas** seleccionamos 8 (para que se vea el nombre completo de la variable).
    - En **Medida** seleccionamos **Ordinal**
  - b) En la carpeta **Vista de variables**, se selecciona el número de fila de la variable **opinión1**, con lo que la fila entera quedará iluminada.
  - c) Se selecciona la opción **Edición** ⇒ **Copiar**
  - d) Se hace *clic* en el número de la fila vacía (en color gris) situada bajo **opinión1**
  - e) Se selecciona la opción **Edición** ⇒ **Pegar**
  - f) Se cambia el **Nombre** de esta nueva variable que se denominará **opinión2**. Se modifica un poco la **Etiqueta** de esta nueva variable que ahora será: *opinión del segundo experto respecto de la encuadernación del libro*.
  - g) Lo que se ha hecho para crear la variable **opinión2** se hace ahora para crear la nueva variable **opinión3** cuya etiqueta se modificará de la siguiente forma: *opinión del tercer experto respecto de la encuadernación del libro*.
  - h) Volver a grabar todo en **Libros.sav**

El resultado de todo el proceso anterior en la carpeta **Vista de variables** es lo que aparece en la Figura 9.

Para terminar este apartado, recordaremos que en cualquier momento se pueden cambiar las especificaciones declaradas en la definición de cualquier variable. 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.

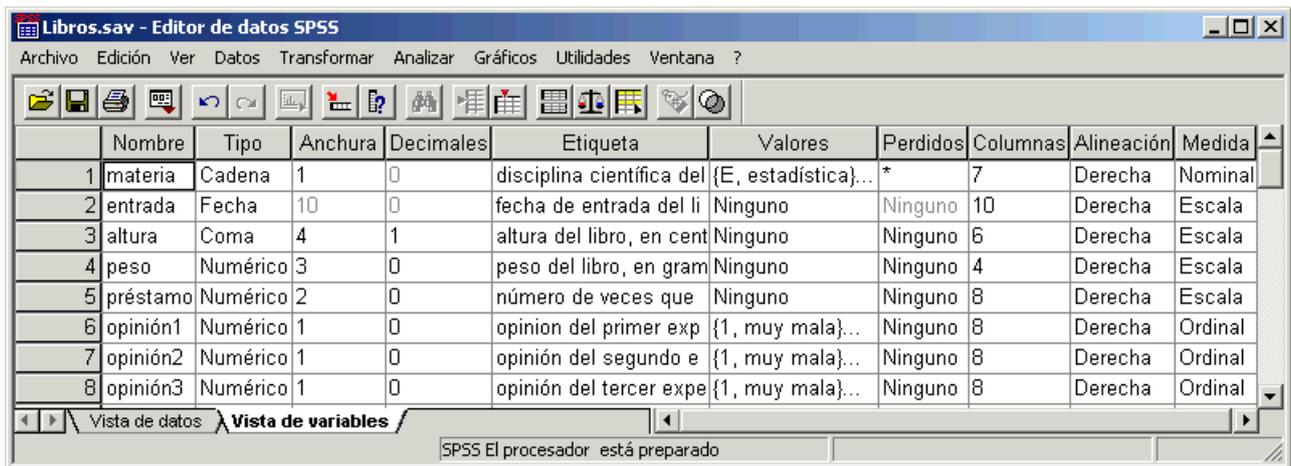


Figura 9: Ventana del editor de datos con las variables definidas

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

En la carpeta **Vista de datos** de la ventana del *editor de datos* se observa que las variables definidas están activas, aunque las casillas correspondientes a sus datos o resultados están vacías.

Antes de introducir los datos, vamos a leer lo siguiente:

1. Recomiendo introducir los datos por columnas (variables), para ir grabando los datos en **Libros.sav** cada vez que introduzcamos una nueva columna (variable).
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. La fecha de **entrada** de los libros 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 03.05.1995, pues es la que hemos elegido en **Tipo**. 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.

Recomiendo introducir las fechas con el separador /, pues éste está en el teclado numérico (por ejemplo, teclearemos 3/5/95).

Una vez hechas estas consideraciones introducimos nuestros datos, menos los resultados de la variable **precio**, que añadiremos posteriormente. Para hacer la introducción de datos a través del teclado tenemos que volver a la Figura 2 donde están los resultados de las variables que hemos definido. Al final debemos tener la matriz de datos de la Figura 10. No olvidemos grabar ahora el archivo de datos en **Libros.sav**. 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 ).

## 1.8. Lectura de archivos de datos propios de SPSS

Para leer datos grabados mediante **SPSS** hay que elegir **Archivo** ⇒ **Abrir** ⇒ **Datos** 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** ⇒ **Datos usados recientemente** (ver Figura 11). Si se encuentra aquí el nombre del archivo, se puede también abrir haciendo *clic* sobre dicho nombre. Para practicar, podemos abrir ahora otro archivo de datos, que no sea **Libros.sav**. Si lo hacemos veremos que, además de abrirse el archivo de datos seleccionado, se abre la *ventana del Visor SPSS*. La información que aparece en dicha ventana no nos interesa, por lo que podemos cerrarla (pulsando el botón ). Observamos, entonces, que el sistema nos pregunta si queremos guardar los resultados (¿*Desea guardar el contenido del Visor de resultados en Resultados1?*), a lo que nosotros responderemos que *No*. (En el apartado 2.2 daremos más información sobre la *ventana del Visor SPSS*).

## 1.9. Edición y modificación de los datos

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

	materia	entrada	altura	peso	préstamo	opinión1	opinión2	opinión3
1	E	03.05.1999	20.5	325	32	2	1	1
2	G	06.12.2000	25.6	890	18	5	4	5
3	E	25.10.2000	20.0	415	9	4	5	5
4	I	03.09.2001	20.0	400	15	2	3	4
5	L	12.03.2000	18.8	515	2	3	3	3
6	I	26.06.2002	25.0	650	36	1	5	3
7	H	15.10.1999	30.6	790	5	3	3	1
8	H	26.11.2001	28.9	890	9	1	3	2
9	L	07.07.2000	20.1	320	7	4	2	1
10	I	04.05.2002	20.0	420	12	2	5	3
11	E	06.10.2000	23.5	620	42	3	2	2
12	E	12.12.2001	23.5	720	20	4	3	4
13	G	12.05.2002	30.9	850	10	3	3	5
14	I	13.11.1999	20.5	620	9	1	3	2
15	E	15.02.2002	22.0	475	15	2	5	5

Figura 10: Ventana del editor rellena con los datos del ejemplo (menos la variable precio)

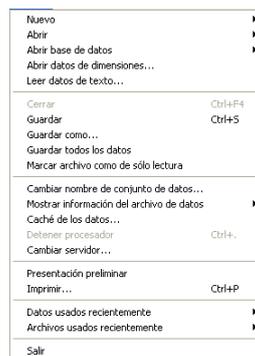


Figura 11: Menú Archivo

- 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  $\leftrightarrow$  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 **Edición**  $\Rightarrow$  **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 *clic* sobre el número del caso. Para seleccionar una variable (columna) se hace *clic* sobre el nombre de ella. Para seleccionar un rectángulo de datos se hace *clic* 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**  $\Rightarrow$  **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 casos (libros) numerados con el 6 y el 22.

- Para **insertar un nuevo caso** entre otros ya existentes, hacemos *clic* sobre el número del caso siguiente (para seleccionarlo) y elegimos **Edición**  $\Rightarrow$  **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 la opción **Insertar casos** en el menú contextual que aparece.

Si queremos insertar nuevos casos al final de los ya existentes, basta con introducirlos de la forma indicada en el apartado 1.7.

- Para **ir a una variable** concreta se elige la opción **Utilidades** ⇒ **Variables** y en el cuadro de diálogo que aparece (Figura 12) se selecciona la variable deseada y se pulsa el botón **Ir a**. También aparece este cuadro de diálogo pulsando el botón del octavo icono de la barra de herramientas.

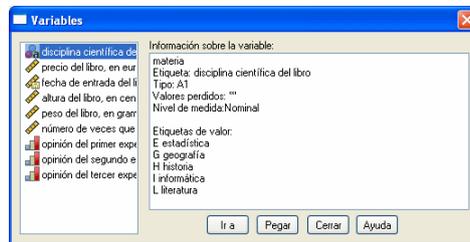


Figura 12: Cuadro que aparece con Utilidades ⇒ Variables

- Para **insertar una variable** entre otras ya existentes, hacemos *clic* sobre el nombre de la variable siguiente (para seleccionarla) y elegimos **Edición** ⇒ **Insertar variable**, o bien empleamos el undécimo icono de la barra de herramientas, o bien pulsamos el botón auxiliar del ratón y elegimos la opción **Insertar variables** en el menú contextual que aparece. Esto se puede hacer tanto en la carpeta **Vista de datos** como en la carpeta **Vista de variables**.

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 1.4 y 1.7.

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

- En la carpeta **Vista de variables** seleccionamos a variable **entrada** haciendo *clic* sobre su nombre.
- Elegimos la opción **Edición** ⇒ **Insertar variable**.
- Una nueva variable es generada, con un nombre genérico (**var00001**), y con el tipo y formato preestablecido por el sistema (ver Figura 13).

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	materia	Cadena	1	0	disciplina cient	{E, estadística *		7	Derecha	Nominal
2	var00001	Numérico	8	2		Ninguno	Ninguno	8	Derecha	Escala
3	entrada	Fecha	10	0	fecha de entra	Ninguno	Ninguno	10	Derecha	Escala
4	altura	Coma	4	1	altura del libro,	Ninguno	Ninguno	6	Derecha	Escala
5	peso	Numérico	3	0	peso del libro,	Ninguno	Ninguno	4	Derecha	Escala
6	préstamo	Numérico	2	0	número de vec	Ninguno	Ninguno	8	Derecha	Escala
7	opinión1	Numérico	1	0	opinión del pri	{1, muy mala}	Ninguno	8	Derecha	Ordinal
8	opinión2	Numérico	1	0	opinión del seg	{1, muy mala}	Ninguno	8	Derecha	Ordinal
9	opinión3	Numérico	1	0	opinión del ter	{1, muy mala}	Ninguno	8	Derecha	Ordinal

Figura 13: Carpeta Vista de Variables con una nueva variable insertada

En la carpeta **Vista de datos** se puede observar que la columna correspondiente tiene 30 datos ausentes (ver Figura 14).

- En la carpeta **Vista de variables** definimos ahora la variable (ver el apartado 1.4) con las siguientes especificaciones:
  - **Nombre** (de variable): **precio**
  - **Tipo**: *Numérica*, **Anchura**: 3, **Cifras decimales**: 0
  - **Etiqueta** (de variable): *precio del libro, en euros*
  - **Columnas**: 6
- En la carpeta **Vista de datos** introducimos ahora los resultados de esta variable de la forma indicada en el apartado 1.7. Dichos resultados se encuentran, como ya sabemos, en la Figura 2.

	materia	var00001	entrada	altura	peso	préstamo	opinión1	opinión2	opinión3
1	E		03.05.1999	20.5	325	32	2	1	1
2	G		06.12.2000	25.6	890	18	5	4	5
3	E		25.10.2000	20.0	415	9	4	5	5
4	I		03.09.2001	20.0	400	15	2	3	4
5	L		12.03.2000	18.8	515	2	3	3	3
6	I		26.06.2002	25.0	650	36	1	5	3
7	H		15.10.1999	30.6	790	5	3	3	1
8	H		26.11.2001	28.9	890	9	1	3	2
9	L		07.07.2000	20.1	320	7	4	2	1
10	I		04.05.2002	20.0	420	12	2	5	3

Figura 14: Carpeta Vista de datos con una nueva variable insertada

- Para **cambiar las especificaciones de una variable** se procede de forma análoga a la definición inicial de las variables (ver los apartados 1.4 y 1.6). 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 12. 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**. *Deshacer* también está en el quinto icono de la barra de herramientas.

### 1.10. Modificación de la apariencia del editor de datos

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 *click* en la opción deseada del menú que aparece (Figura 15). Estas opciones son:



Figura 15: 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*.

**Variables (o Datos):** Sirve para ver la carpeta **Vista de variables** o la carpeta **Vista de datos**.

Por ejemplo, en la Figura 16 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**.

	materia	precio	entrada	altura	peso	préstamo	opinión1	opinión2	opinión3
1	estadística	110	03.05.1999	20.5	325	32	mala	muy mala	muy mala
2	geografía	30	06.12.2000	25.6	890	18	muy buena	buena	muy buena
3	estadística	75	25.10.2000	20.0	415	9	buena	muy buena	muy buena
4	informática	45	03.09.2001	20.0	400	15	mala	regular	buena
5	literatura	32	12.03.2000	18.8	515	2	regular	regular	regular
6	informática	69	26.06.2002	25.0	650	36	muy mala	muy buena	regular

Figura 16: Editor de datos mostrando las etiquetas de las variables (y sin la cuadrícula)

## 1.11. Impresión de los datos

Para imprimir los datos hay que seleccionar la opción **Archivo** ⇒ **Imprimir** o usar el tercer icono de la barra de herramientas.

## 1.12. Importación de archivos de datos

Acabamos de ver en los apartados anteriores los procedimientos que se emplean para generar los archivos de datos propios de *SPSS para Windows*. En el apartado 1.8 vimos como leer (importar) este tipo de archivos de datos. Pero en algunas ocasiones podemos disponer de datos ya grabados en un archivo generado por una aplicación informática distinta de *SPSS*. Si se desea analizar estadísticamente tal información no es preciso crear un archivo *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 archivos (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** ⇒ **Datos** y en el cuadro de diálogo que aparece se indica *Systat(\*.sys)* en *Tipo*).
- Archivos de **hojas de cálculo**, como *Lotus*, *Excel* o *formato SYLK* (se selecciona **Archivo** ⇒ **Abrir** ⇒ **Datos** y en *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** ⇒ **Datos** y en el cuadro de diálogo que aparece se indica *dBASE(\*.dbf)* en *Tipo*).
- Por supuesto, archivos de datos *SPSS* creados en otros entornos y sistemas operativos (se selecciona **Archivo** ⇒ **Abrir** ⇒ **Datos** y en *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 archivos portátiles creados en otros entornos como Macintosh o VMS).
- Además, se pueden capturar y manejar datos de **Access**, **Excel** o **FoxPro** de las últimas versiones (se selecciona **Archivo** ⇒ **Abrir base de datos** ⇒ **Nueva consulta**).

## 1.13. Transformaciones en el archivo de datos

Con esta versión de *SPSS* cada vez que hagamos una transformación en el archivo de datos se abrirá la *ventana del Visor SPSS*. En general, la información que aparecerá en dicha ventana no nos interesará por lo que la cerraremos (pulsando el botón ). Observaremos que el sistema nos preguntará si queremos guardar los resultados (*¿Desea guardar el contenido del Visor de resultados en Resultados?*) a lo que nosotros responderemos que *No*.

### 1.13.1. Transformación de variables sin utilizar el condicional

En este apartado y en el siguiente vamos a ver el modo de generar nuevas variables mediante transformaciones efectuadas sobre los valores de las variables ya definidas.

Siempre que realicemos transformaciones en nuestro archivo **Libros.sav** grabaremos estos nuevos datos en un archivo nuevo que se denominará **Libros2.sav**

En el Cuadro 1 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

()	Paréntesis
**	Exponenciación
*	Multiplicación
/	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 1: Operaciones aritméticas, relacionales y lógicas que se pueden realizar con SPSS

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 el Cuadro 1 (de arriba a abajo).

Para construir una nueva variable mediante transformaciones de otras ya existentes, se tiene que elegir la opción **Transformar** ⇒ **Calcular variable** con lo que se abre el cuadro de diálogo de la Figura 17. En esta ventana tenemos cinco

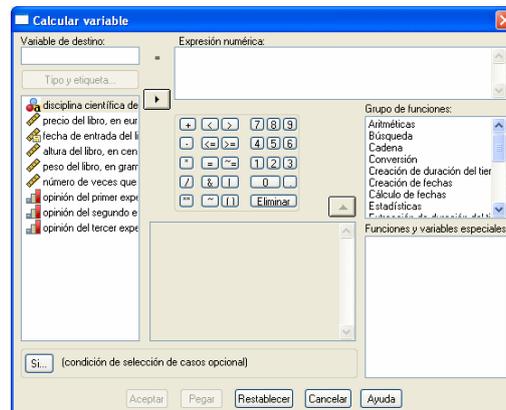


Figura 17: Cuadro de diálogo para transformar variables

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 a su derecha está el grupo de funciones disponibles (**Grupo de funciones:**). Una vez seleccionado un grupo dentro de la lista **Grupo de funciones:**, debajo aparece otra lista con las funciones y variables que se pueden usar dentro del grupo de funciones elegido (**Funciones y variables especiales:**).

Vamos a leer, a continuación, la forma de hacer una transformación de variables. Después practicaremos esta opción.

- 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 los apartados 1.4 y 1.6.
- 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 archivo 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.

Una de las variables del archivo **Libros.sav** es la *altura* de cada libro, medida en centímetros, cuyos resultados vienen expresados con una cifra decimal. Como ejemplo, vamos a expresar la *altura* en milímetros; es decir, vamos a multiplicar todos los resultados de dicha variable por 10. Para ello, seleccionamos **Transformar** ⇒ **Calcular variable**. Como en este caso

no es necesario generar una nueva variable, en el recuadro **Variable de destino** se escribe el mismo nombre (**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; a lo que nosotros contestaremos que *Sí*.

Lo conveniente ahora es cambiar la definición de la variable **altura** (ver los apartados 1.4 y 1.6), poniendo **Tipo=Numérica**, **Anchura=3**, **Cifras decimales=0** y **Etiqueta=altura del libro, en milímetros**. No olvidemos ahora grabar estos nuevos datos en un archivo con el nombre **Libros2.sav**

Hecho esto, en la ventana del *editor de datos* tendríamos lo que aparece en la Figura 18.

	materia	precio	entrada	altura	peso	préstamo	opinión1	opinión2	opinión3
1	E	110	03.05.1999	205	325	32	2	1	1
2	G	30	06.12.2000	256	890	18	5	4	5
3	E	75	25.10.2000	200	415	9	4	5	5
4	I	45	03.09.2001	200	400	15	2	3	4
5	L	32	12.03.2000	188	515	2	3	3	3

Figura 18: Ventana del editor con la nueva variable altura

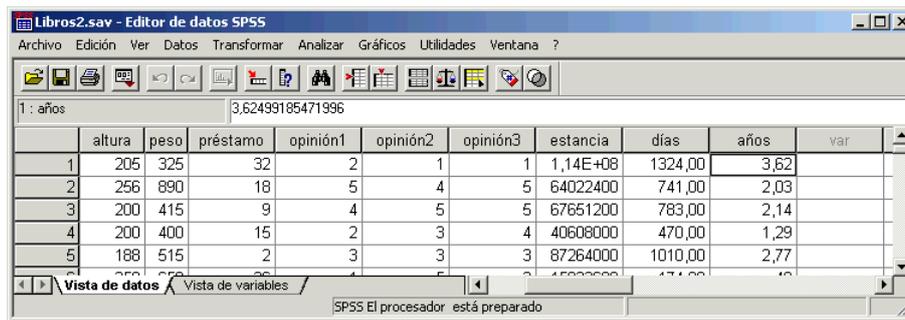
Siguiendo ahora con el archivo **Libros2.sav**, en él consta una variable de tipo fecha, que es la fecha de entrada del libro a la biblioteca. Si quisiéramos saber el tiempo transcurrido entre esta fecha y la fecha 17/12/2002 podríamos crear una nueva variable, a la que podríamos llamar **estancia** con el resultado de la diferencia entre la fecha 17/12/2002 y la fecha que aparece en la variable **entrada**.

Para ello, seleccionamos **Transformar** ⇒ **Calcular variable**. En **Variable de destino** escribimos **estancia** (ver Figura 17). Ahora tenemos que expresar la diferencia entre la fecha 17/12/2002 y las fechas de la variable **entrada**. Pero la fecha 17/12/2002 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. Para ello, en **Grupo de funciones** hacemos *clic* sobre **Creación de fechas**; en la lista de **Funciones y variables específicas** hacemos doble *clic* sobre **Date.Dmy**. Podemos observar que en el recuadro que hay a la izquierda de la lista **Funciones y variables específicas** aparece una explicación de lo que hace concretamente la función seleccionada. En **Expresión numérica** aparece lo siguiente:  $DATE.DMY(?,?,?)$ . Tenemos que sustituir las interrogaciones por la fecha que queremos; es decir, debe aparecer  $DATE.DMY(17,12,2002)$ . Luego pulsamos en el signo menos de la calculadora (–) y después nos vamos a la lista de variables y hacemos doble *clic* sobre la variable *fecha de entrada* .... Por tanto, en el recuadro **Expresión numérica** nos aparecerá:  $DATE.DMY(17,12,2002)-entrada$ . Después de pulsar el botón **Aceptar** obtendremos en la carpeta de datos, a continuación de nuestra última variable (**opinión3**), los resultados de la nueva variable **estancia** (ver Figura 19). Podemos volver a grabar estos nuevos datos en **Libros2.sav**

Como vemos, los valores de la variable **estancia** 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 ello, creamos una nueva variable, que llamaremos **días**, que será el resultado de la siguiente **Expresión numérica**:  $estancia/86400$  (que es recomendable poner con ayuda de la lista de variables y de la calculadora del cuadro de diálogo), pues cada día tiene 86400 segundos. Después de pulsar el botón **Aceptar** obtendremos en la carpeta de datos, a continuación de nuestra última variable, los resultados de la nueva variable **días** (ver Figura 19). Podemos volver a grabar estos nuevos datos en **Libros2.sav**

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**:  $dias/365.2422$  (que es recomendable poner con ayuda de la lista de variables y de la calculadora del cuadro de diálogo), pues el año tiene 365'2422 días. Los valores de esta nueva variable aparecen en la columna **años** de la Figura 19. 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 carpeta *vista de datos* aparezcan más de dos decimales tenemos que modificar las **Cifras decimales** de la variable **años** señalando el número de decimales deseado. Hagamos o no esta modificación, 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** lo siguiente: *años de estancia del libro en la biblioteca*.

Aunque lo he hecho por su interés pedagógico, no es lógico conservar las variables **estancia** y **días**, por lo que podemos



Libros2.sav - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : años 3,62499185471996

	altura	peso	préstamo	opinión1	opinión2	opinión3	estancia	días	años	var
1	205	325	32	2	1	1	1,14E+08	1324,00	3,62	
2	256	890	18	5	4	5	64022400	741,00	2,03	
3	200	415	9	4	5	5	67651200	783,00	2,14	
4	200	400	15	2	3	4	40608000	470,00	1,29	
5	188	515	2	3	3	3	87264000	1010,00	2,77	

Vista de datos Vista de variables

SPSS El procesador está preparado

Figura 19: Ventana del editor con las nuevas variables estancia, días y años

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 **Borrar**.

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 20) insertamos primero una variable vacía (que automáticamente se llamará **var0001**) entre las variables **entrada** y **altura** (hacemos *click* sobre el nombre de la variable **altura** y elegimos **Edición** ⇒ **Insertar variable**); luego seleccionamos la variable **años**, la *cortamos*, seleccionamos **var0001** y *pegamos* (para hacer este proceso, se puede consultar el apartado 1.9). Ahora podemos volver a grabar los datos en **Libros2.sav**. Para ver todas las funciones de fecha y hora disponibles se puede consultar la Ayuda de **SPSS**.



Libros2.sav - Editor de datos SPSS

Archivo Edición Ver Datos Transformar Analizar Gráficos Utilidades Ventana ?

1 : años 3,62499185471996

	materia	precio	entrada	años	altura	peso	préstamo	opinión1	opinión2	opinión3
1	E	110	03.05.1999	3,62	205	325	32	2	1	1
2	G	30	06.12.2000	2,03	256	890	18	5	4	5
3	E	75	25.10.2000	2,14	200	415	9	4	5	5
4	I	45	03.09.2001	1,29	200	400	15	2	3	4
5	L	32	12.03.2000	2,77	188	515	2	3	3	3

Vista de datos Vista de variables

SPSS El procesador está preparado

Figura 20: Ventana del editor con la variable años junto a la variable entrada

### 1.13.2. Transformación condicional de variables

En el apartado anterior hemos descrito las posibilidades que brinda el cuadro de diálogo correspondiente a la opción **Transformar** ⇒ **Calcular variable**; pero hemos hecho caso omiso a un botón: el de **Si...** (*condición de selección de casos opcional*), 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.

Para tal fin, una vez seleccionada la opción **Transformar** ⇒ **Calcular variable**, debe pulsarse el botón **Si...** (*condición de selección de casos opcional*) con lo que se abre el cuadro de la Figura 21. 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 de la **altura** por el **peso** sólo para los libros de *estadística*, y quisiéramos llamar a la nueva variable **robustez**. Para ello, en este cuadro de diálogo de la Figura 17, junto a **Variable de destino** escribimos **robustez**; en **Expresión numérica** ponemos  $altura * peso$  (como ya he dicho varias veces, con ayuda de la lista de variables y de la calculadora de este cuadro de diálogo) y pulsamos el botón **Si...** (*condición de selección de casos opcional*). En el cuadro de diálogo resultante activamos **Incluir si el caso satisface la condición**; a continuación ponemos (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

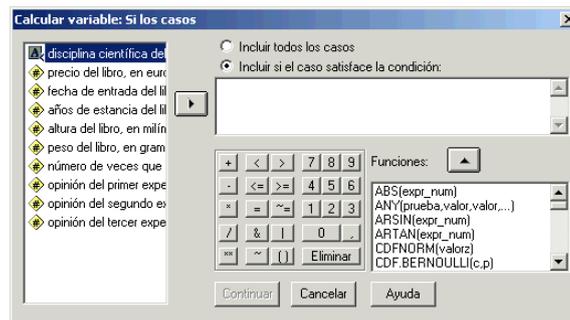


Figura 21: Cuadro de diálogo para hacer la modificación condicional de una variable

interrogación.) Después se pulsa el botón **Continuar** y volvemos al cuadro de diálogo de la Figura 17. En este cuadro de diálogo pulsamos **Aceptar**. Así, 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 22). El resto de los casos están en blanco (son valores perdidos o ausentes).

	materia	precio	entrada	años	altura	peso	préstamo	opinión1	opinión2	opinión3	robustez
1	E	110	03.05.1999	3,62	205	325	32	2	1	1	66625
2	G	30	06.12.2000	2,03	256	890	18	5	4	5	
3	E	75	25.10.2000	2,14	200	415	9	4	5	5	83000
4	I	45	03.09.2001	1,29	200	400	15	2	3	4	
5	L	32	12.03.2000	2,77	188	515	2	3	3	3	
6	I	69	26.06.2002	,48	250	650	36	1	5	3	
7	H	30	15.10.1999	3,17	306	790	5	3	3	1	
8	H	34	26.11.2001	1,06	289	890	9	1	3	2	
9	L	42	07.07.2000	2,44	201	320	7	4	2	1	
10	I	46	04.05.2002	,62	200	420	12	2	5	3	
11	E	53	06.10.2000	2,20	235	620	42	3	2	2	145700

Figura 22: Ventana del editor con la nueva variable robustez

Sería conveniente poner una etiqueta explicativa a la variable **robustez**. Para ello, en la carpeta **Vista de variables**, en **Etiqueta** escribimos: *producto de la altura por el peso, sólo para los libros de estadística*. Después de estos cambios, volvemos a grabar el archivo de datos **Libros2.sav**

### 1.13.3. Clasificación de variables mediante la opción **Transformar** ⇒ **Recodificar ...**

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: se puede recodificar guardando el nuevo plan de codificación en la misma variable, con lo que se perderán los antiguos valores (opción **Transformar** ⇒ **Recodificar en las mismas variables**) o 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 manera que el archivo va a contar con dos variables que representan el mismo fenómeno pero con dos planes de codificación distintos (opción **Transformar** ⇒ **Recodificar en distintas variables**). Nosotros vamos a ver un ejemplo de este segundo caso.

Supongamos que en nuestro archivo **Libros2.sav** queremos clasificar los datos de la variable **préstamo** según lo indicado en el Cuadro 2.

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 23. En este cuadro de diálogo seleccionamos la variable que queremos recodificar (haciendo doble *click* sobre ella en la lista de variables que hay a la izquierda, o haciendo un *click* y luego

valor antiguo (de la variable <b>préstamo</b> )	valor nuevo
desde el menor hasta 10	1
entre 11 y 20	2
entre 21 y 30	3
entre 31 y el mayor valor	4

Cuadro 2: Clasificación de los resultados de la variable préstamo

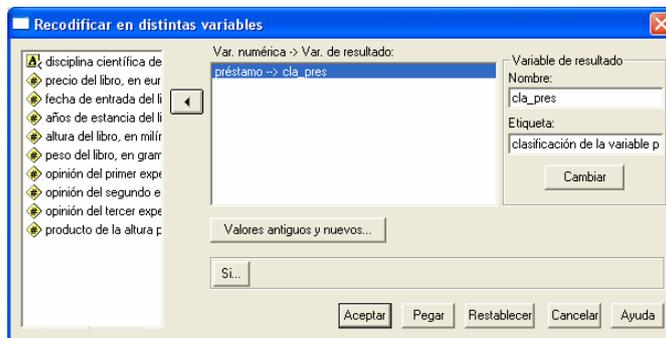


Figura 23: Cuadro de diálogo para recodificar una variable

pulsando el botón ) , es decir, seleccionamos la variable **préstamo**. En **Variable de resultado** escribimos **cla\_pres** debajo de **Nombre** y **clasificación de la variable préstamo en cuatro categorías** 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 24. En este cuadro

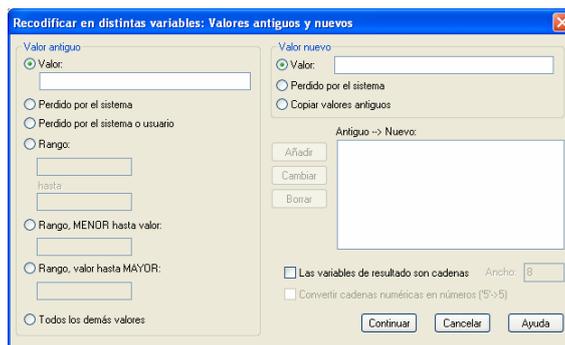


Figura 24: Cuadro de diálogo para hacer la asignación de nuevos códigos

tenemos que hacer la asignación deseada de forma muy similar a como hicimos la asignación de etiquetas de valor de una variable:

- En **Valor antiguo** seleccionamos **Rango: MENOR hasta valor** y en éste escribimos 10 (esto significa desde el menor valor de la variable hasta 10). En **Valor nuevo**, junto a **Valor** escribimos 1. Luego pulsamos el botón **Añadir**, con lo que este cambio queda reflejado en el cuadro que hay bajo **Antiguo → Nuevo**.
- En **Valor antiguo** seleccionamos **Rango:** y escribimos 11 **hasta** 20. En **Valor nuevo**, junto a **Valor** escribimos 2. A continuación pulsamos el botón **Añadir**.
- En **Valor antiguo** seleccionamos **Rango:** y escribimos 21 **hasta** 30. En **Valor nuevo**, junto a **Valor** escribimos 3. A continuación pulsamos el botón **Añadir**.
- En **Valor antiguo** seleccionamos **Rango: valor hasta MAYOR** y escribimos 31 (esto quiere decir desde 31 hasta el mayor). En **Valor nuevo**, junto a **Valor** escribimos 4. 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 23. En este cuadro pulsamos el botón **Aceptar**. Como consecuencia, en el *editor de datos* tenemos una columna más con la variable **cla\_pres**. A continuación, en la carpeta **Vista de variables** debemos cambiar la definición de la nueva variable **cla\_pres** de la siguiente manera: en **Cifras decimales** seleccionamos un 0, en **Anchura** seleccionamos un 1, en **Medida** seleccionamos **Ordinal** y en **Valores** ponemos las etiquetas (o explicaciones) de los nuevos resultados:

Hacemos *click* en la parte derecha de la casilla intersección de la fila correspondiente a nuestra variable con la columna correspondiente a **Valores**. En el cuadro de diálogo que aparece (Figura 5) hacemos lo siguiente:

- Junto a **Valor** escribimos 1, junto a **Etiqueta** escribimos  $\leq 10$  y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos 2, junto a **Etiqueta** escribimos 11 – 20 y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos 3, junto a **Etiqueta** escribimos 21 – 30 y pulsamos el botón **Añadir**;
- Junto a **Valor** escribimos 4, junto a **Etiqueta** escribimos  $\geq 31$  y pulsamos el botón **Añadir**.

Una vez finalizado el proceso se pulsa **Aceptar**. El *editor de datos* aparecería como en la Figura 25. Por último, grabamos, de nuevo, el archivo **Libros2.sav**

	materia	precio	entrada	años	altura	peso	préstamo	opinión1	opinión2	opinión3	robustez	cla_pres
1	E	110	03.05.1999	3,62	205	325	32	2	1	1	66625	4
2	G	30	06.12.2000	2,03	256	890	18	5	4	5	.	2
3	E	75	25.10.2000	2,14	200	415	9	4	5	5	83000	1
4	I	45	03.09.2001	1,29	200	400	15	2	3	4	.	2
5	L	32	12.03.2000	2,77	188	515	2	3	3	3	.	1
6	I	69	26.06.2002	,48	250	650	36	1	5	3	.	4

Figura 25: Ventana del editor con la nueva variable cla\_pres

#### 1.13.4. Clasificación de variables mediante la opción *Transformar* $\Rightarrow$ *Agrupación visual*

La opción **Transformar** $\Rightarrow$ **Agrupación visual** sirve para agrupar los datos de las variables cuantitativas en intervalos de clase. Dichas variables pueden ser continuas o discretas. Incluso se puede utilizar para variables ordinales (aunque, en general, no tiene mucho sentido hacerlo).

Cuando se selecciona **Transformar** $\Rightarrow$ **Agrupación visual** aparece el cuadro de diálogo de la Figura 26. En este cuadro de

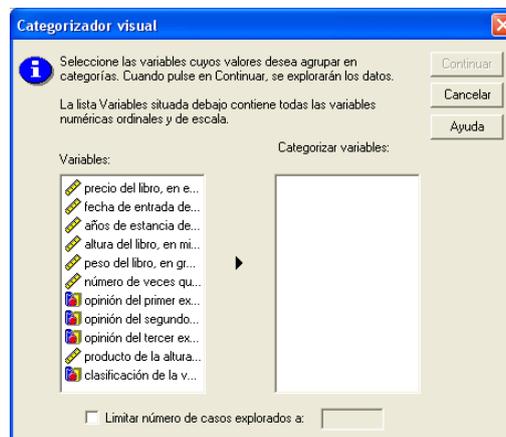


Figura 26: Cuadro de diálogo de la opción *Transformar*  $\Rightarrow$  *Agrupación visual*

diálogo seleccionamos las variables de las cuales queremos agrupar sus resultados en intervalos. Por ejemplo, del archivo **Libros2.sav** seleccionamos la variable **precio**. Para ello, hacemos un *clic* sobre el nombre (o la etiqueta) de la variable **precio** y pulsamos el botón . A continuación pulsamos el botón **Continuar**, con lo que nos aparece el cuadro de diálogo de la Figura 27.

Vamos a agrupar los datos de la variable **precio** en cinco intervalos de la misma amplitud. Para ello hacemos un *clic* sobre su nombre (o su etiqueta) en **Lista de variables exploradas** del cuadro de diálogo de la Figura 27, con lo cual nos aparece, entre otras cosas, un histograma que muestra la distribución de dicha variable (ver la Figura 28). Además, se nos muestra el nombre de la variable actual y su etiqueta; una etiqueta (que podemos cambiar) para la variable clasificada en intervalos; el mínimo valor de la variable (30 euros); el máximo valor (110 euros); se nos indica que hay 30 casos explorados y que no hay ningún valor perdido. Como el recorrido de la variable es  $R = 110 - 30 = 80$  euros entonces la amplitud de cada uno de los cinco intervalos es igual a  $d = 80/5 = 16$  euros.

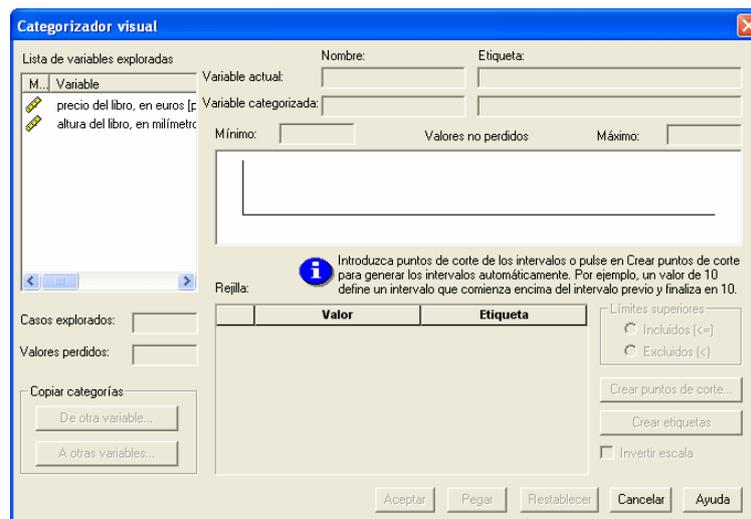


Figura 27: Cuadro de diálogo para clasificar en intervalos

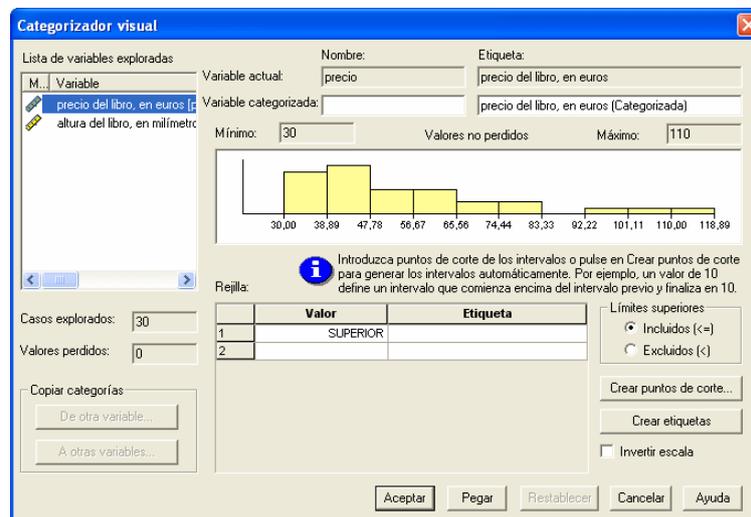


Figura 28: Cuadro de diálogo para hacer la clasificación de la variable precio en intervalos

Lo primero que debemos hacer es poner el nombre a la nueva variable, para lo cual escribimos, por ejemplo, *precio\_c* junto a **Variable agrupada**. Modificamos también la etiqueta de la nueva variable, poniendo en **Etiqueta** de la variable agrupada: *precio (clasificada en 5 intervalos de la misma amplitud)*. A continuación pulsamos en **Crear puntos de corte**, con lo que nos aparece el cuadro de diálogo de la Figura 29.

En este cuadro de diálogo aparecen los tres siguientes apartados:

**Intervalos de igual amplitud:** Genera intervalos de igual amplitud basándose en dos cualesquiera de los tres criterios que aparecen en este apartado.

**Percentiles iguales basados en los casos explorados:** Los datos se clasifican según grupos de percentiles; cada grupo contiene aproximadamente el mismo número de casos.

**Puntos de corte en media y desviaciones típicas seleccionadas, basadas en casos explorados:** Genera clases basándose en los valores de la media ( $\bar{x}$ ) y la cuasi-desviación típica ( $S$ ) de la muestra.

Continuando con el ejemplo iniciado (clasificación de la variable **precio** en cinco intervalos de la misma amplitud), activamos el apartado **Intervalos de igual amplitud**. Los tres sub-apartados los rellenamos de la siguiente manera:

**Posición del primer punto de corte:** Valor que define el límite superior del primer intervalo. En nuestro caso, pondríamos aquí el valor **46**, puesto el mínimo valor es 30 euros y queremos hacer 5 intervalos de una amplitud igual a 16 euros ( $30 + 16 = 46$ ).

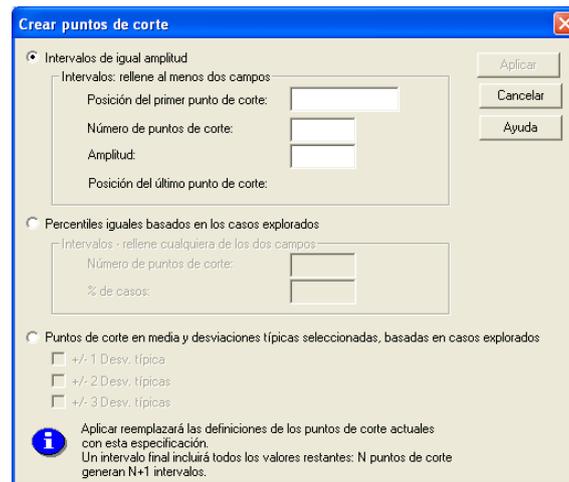


Figura 29: Cuadro de diálogo para crear los intervalos

**Número de puntos de corte:** El número de puntos de corte es igual al número de intervalos menos uno (¡extraño!). Por tanto, aquí nosotros ponemos 4.

**Amplitud:** En nuestro ejemplo no hace falta poner la amplitud pues con los dos criterios anteriores ya se sabe que la amplitud es igual a 16. Pero también podemos escribirlo.

En el cuadro de diálogo de la Figura 29, se activa (otra vez) la opción **Intervalos de la misma amplitud** y después se hace clic sobre **Aplicar**, lo que nos devuelve al cuadro de diálogo de la Figura 30.

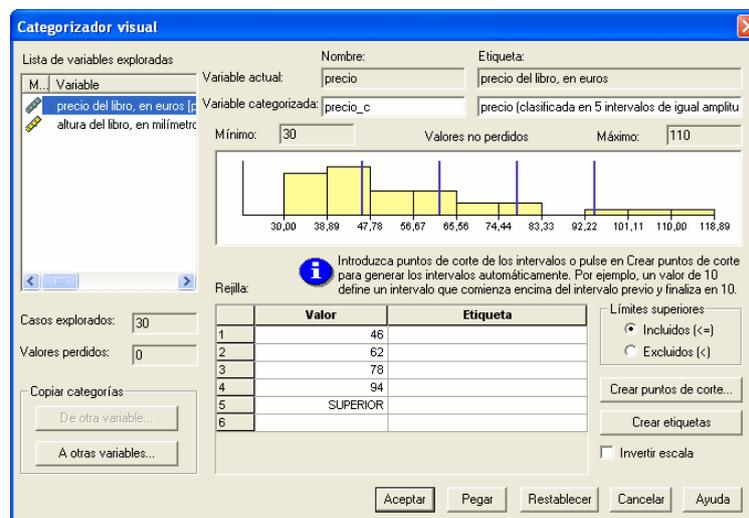


Figura 30: Clasificación de la variable precio en 5 intervalos de la misma amplitud

Ahora podemos pulsar en **Crear etiquetas** para generar automáticamente etiquetas de valor para cada categoría (intervalo). También se pueden introducir o modificar manualmente los puntos de corte y las etiquetas en la **Rejilla**. Además, se puede cambiar la ubicación de los puntos de corte, arrastrando las líneas verticales que los marcan en el histograma. Para acabar el proceso de agrupación en intervalos, se pulsa en **Aceptar**.

Puesto que la nueva variable se añade al final del archivo de datos, aparece en la columna que se encuentra más a la derecha en la **Vista de datos** (ver la Figura 31) y en la última fila de la **Vista de variables** (ver la Figura 32). No olvidemos grabar, otra vez, el archivo **Libros2.sav**

Vamos ahora a agrupar los datos de la variable **altura** del archivo **Libros2.sav** en 4 intervalos utilizando la opción **Percentiles iguales basados en los casos explorados** en **Crear puntos de corte** (ver la Figura 29). Recordemos que la variable **altura** del archivo **Libros2.sav** da los resultados de la altura de los libros, en milímetros. A la nueva variable le pondremos el nombre **altura\_c**.

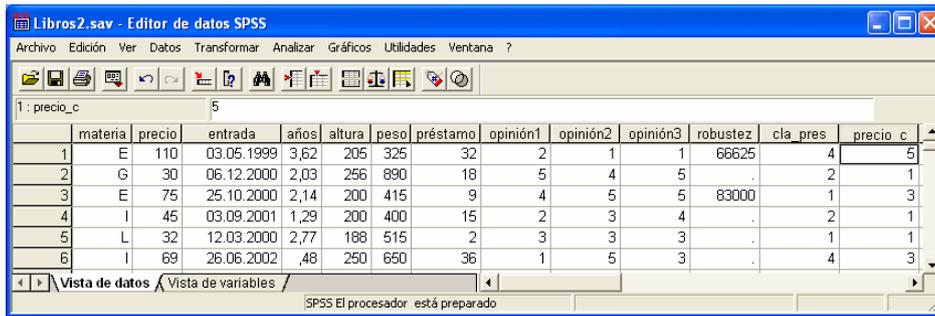


Figura 31: Vista de datos con la nueva variable precio\_c



Figura 32: Vista de variables con la nueva variable precio\_c

Para hacer dicha agrupación, seleccionamos **Transformar**⇒**Agrupación visual**. En el cuadro de diálogo que aparece (Figura 26) seleccionamos la variable **altura** y pulsamos el botón **Continuar**. En el nuevo cuadro de diálogo (Figura 27) hacemos un clic sobre el nombre (o la etiqueta) de la variable **altura** en la **Lista de variables exploradas** (ver la Figura 33). A continuación

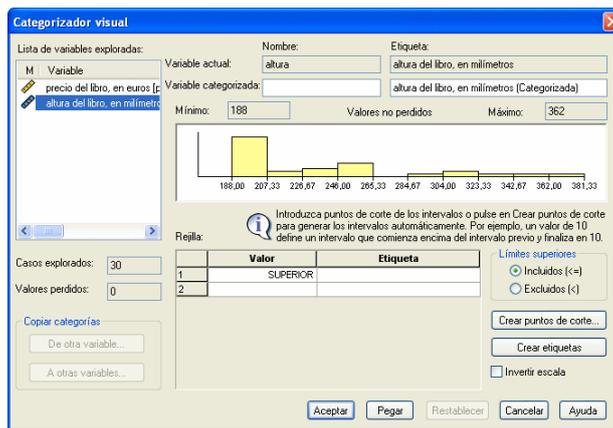


Figura 33: Cuadro de diálogo para hacer la clasificación de la variable altura en 4 intervalos

escribimos **altura\_c** junto a **Variable agrupada**. Modificamos también la etiqueta de la nueva variable, poniendo en **Etiqueta** de la variable agrupada: *altura (clasificada en 4 intervalos basados en los percentiles)*. Después pulsamos en **Crear puntos de corte**, con lo que nos aparece el cuadro de diálogo de la Figura 29. En este cuadro de diálogo, activamos la opción **Percentiles iguales basados en los casos explorados**. Con esta opción los datos se clasifican según grupos de percentiles; cada grupo contiene aproximadamente el mismo número de casos. Por ejemplo, una especificación de cuatro categorías asignaría el valor 1 a los casos situados bajo el cuantil al 25 % (primer cuartil), 2 a los casos situados entre el cuantil al 25 % y el cuantil al 50 % (segundo cuartil o mediana), 3 a los casos situados entre el cuantil al 50 % y el cuantil al 75 % (tercer cuartil) y 4 a los casos situados por encima del cuantil 75 %. Análogamente, una especificación de tres categorías asignaría el valor 1 a los casos situados bajo el cuantil al 33' 3 % (un tercio del total de datos), 2 a los casos situados entre el cuantil al 33' 3 %

y el cuantil al  $66\frac{2}{3}\%$  (dos tercios del total de datos) y 3 a los casos situados por encima del cuantil  $66\frac{2}{3}\%$ . Se elige uno de los dos criterios siguientes:

**Número de puntos de corte:** El número de puntos de corte es igual al número de intervalos menos uno. Por ejemplo, tres puntos de corte generan cuatro categorías conteniendo cada una de ellas el 25 % de los casos.

**% de casos:** Es la amplitud de cada intervalo, expresado en forma de porcentaje sobre el número total de casos. Por ejemplo, el valor  $33\frac{1}{3}$  generaría tres categorías (dos puntos de corte) conteniendo cada una de ellas el  $33\frac{1}{3}\%$  de los casos.

Continuando con el ejemplo (clasificación de la variable **altura** en cuatro intervalos basados en los percentiles), tecleamos un 3 junto a **Número de puntos de corte** y pulsamos en **Aplicar**. Para acabar el proceso de agrupación en intervalos, se pulsa en **Aceptar**. En la carpeta **Vista de datos** obtendríamos lo que aparece en la Figura 34. Por último, grabamos de nuevo el archivo **Libros2.sav**

	materia	precio	entrada	años	altura	peso	préstamo	opinión1	opinión2	opinión3	robustez	cla_pres	precio_c	altura_c
1	E	110	03.05.1999	3,62	205	325	32	2	1	1	66625	4	5	2
2	G	30	06.12.2000	2,03	256	890	18	5	4	5	.	2	1	3
3	E	75	25.10.2000	2,14	200	415	9	4	5	5	83000	1	3	1
4	I	45	03.09.2001	1,29	200	400	15	2	3	4	.	2	1	1
5	L	32	12.03.2000	2,77	188	515	2	3	3	3	.	1	1	1
6	I	69	26.06.2002	,48	250	650	36	1	5	3	.	4	3	3
7	H	30	15.10.1999	3,17	306	790	5	3	3	1	.	1	1	4
8	H	34	26.11.2001	1,06	289	890	9	1	3	2	.	1	1	4

Figura 34: Vista de datos con la nueva variable altura\_c

En el cuadro de diálogo 29 nos queda por explicar el último criterio para la clasificación de una variable en intervalos: **Puntos de corte en media y desviaciones típicas seleccionadas, basadas en casos explorados**. Como ya se ha visto, este criterio genera categorías basándose en los valores de la media ( $\bar{x}$ ) y la cuasi-desviación típica ( $S$ ) de la muestra:

- Si no se selecciona ninguno de los intervalos de desviación típica, se crearán dos intervalos siendo la media el punto de corte que divida las categorías.
- Se puede seleccionar cualquier combinación de los intervalos de desviación típica, basándose en una, dos o tres desviaciones típicas. Por ejemplo, si se selecciona la opción  $\pm 1$  Desv. Típica se generan cuatro categorías: se asignaría el valor 1 a los casos situados bajo el resultado de  $\bar{x} - S$ , 2 a los casos situados entre  $\bar{x} - S$  y  $\bar{x}$ , 3 a los casos situados entre  $\bar{x}$  y  $\bar{x} + S$  y 4 a los casos situados por encima de  $\bar{x} + S$ .

### 1.13.5. Ordenación de los resultados de una variable

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.

Esto se realiza tras seleccionar **Transformar**  $\Rightarrow$  **Recodificación automática**. Al hacerlo se abre el cuadro de diálogo de la Figura 35.

Supongamos que, del archivo **Libros2.sav**, queremos ordenar los resultados de la variable **préstamo** en orden ascendente. En el cuadro de diálogo de la Figura 35 ponemos **préstamo** en el recuadro **Variable**  $\rightarrow$  **Nuevo nombre** haciendo un clic en el nombre de la variable (o su etiqueta) y luego pulsando el botón . A continuación escribimos junto a **Nombre nuevo** el nombre deseado, por ejemplo **ord\_pres**, y pulsamos en **Añadir nuevo nombre**. Después seleccionamos **Recodificar empezando por: Menor valor**. En el cuadro de diálogo hay otros apartados:

**Usar el mismo esquema de recodificación para todas las variables:** Esta opción le permite aplicar un único esquema de recodificación para todas las variables seleccionadas (si hay más de una), lo que genera un esquema de codificación coherente para todas las variables nuevas.

Si se selecciona esta opción, se aplican las siguientes reglas y limitaciones:

- Todas las variables deben ser del mismo tipo (numéricas o de cadena).

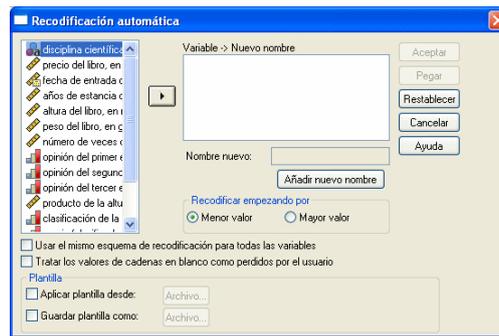


Figura 35: Cuadro de diálogo para la recodificación automática (ordenación de los datos)

- Todos los valores observados para todas las variables seleccionadas se utilizan para crear un orden de valores para recodificar en enteros consecutivos.
- Los valores perdidos definidos por el usuario para las variables nuevas se basan en la primera variable de la lista con valores perdidos definidos por el usuario. El resto de los valores de las demás variables originales, excepto los valores perdidos del sistema, se consideran válidos.

**Tratar los valores de cadena en blanco como valores perdidos definidos por el usuario:** En el caso de las variables de cadena, los valores en blanco o nulos no son tratados como valores perdidos del sistema. Esta opción recodifica automáticamente las cadenas en blanco en un valor perdido definido por el usuario mayor que el valor no perdido máximo.

En nuestro ejemplo, como solamente hemos seleccionado una variable y además dicha variable es numérica, dejamos desactivadas las anteriores opciones y pulsamos **Aceptar**.

El resultado es doble: por un lado, se abre la *ventana del Visor SPSS* con los valores de la variable **préstamo** ordenados de menor a mayor (Figura 36). Podemos observar que la variable **préstamo** tiene 19 resultados distintos. El menor de ellos es el 1 (que ocupa, por tanto, el lugar número 1 en la ordenación) y el mayor de ellos es el 42 (que ocupa, por tanto, el lugar número 19 en la ordenación).

En esta ocasión no nos interesa conservar la información de la *ventana del Visor SPSS*, puesto que ésta quedará reflejada en la carpeta **Vista de datos** del *editor de datos*. Por tanto, podemos cerrar la *ventana del Visor SPSS* y podemos responder *No* a la pregunta (*¿Desea guardar el contenido del Visor de resultados en Resultados?*).

PRÉSTAMO	ORD_PRES	número de veces que se ha prestado
Old Value	New Value	Value Label
1	1	1
2	2	2
4	3	4
5	4	5
7	5	7
8	6	8
9	7	9
10	8	10
11	9	11
12	10	12
13	11	13
15	12	15
16	13	16
18	14	18
20	15	20
23	16	23
32	17	32
36	18	36
42	19	42

Figura 36: Ventana del Visor SPSS con los datos de la variable préstamo ordenados

Por otro lado, en el *editor de datos* aparece la nueva variable **ord\_pres** (Figura 37). En la carpeta **Vista de variables** podemos observar que **SPSS** asigna automáticamente una *Etiqueta* a la nueva variable **ord\_pres** (que coincide con la etiqueta de la variable **préstamo**), así como una *Anchura* (igual a 2) y un nivel de *Medida* (Ordinal). Si queremos, podemos cambiar alguna de estas características de la nueva variable. Como la etiqueta de la nueva variable **ord\_pres** es la misma que la de la variable **préstamo**, es conveniente cambiarla; por ejemplo, podemos ponerle como nueva etiqueta: *ordenación de la variable préstamo de menor a mayor*.

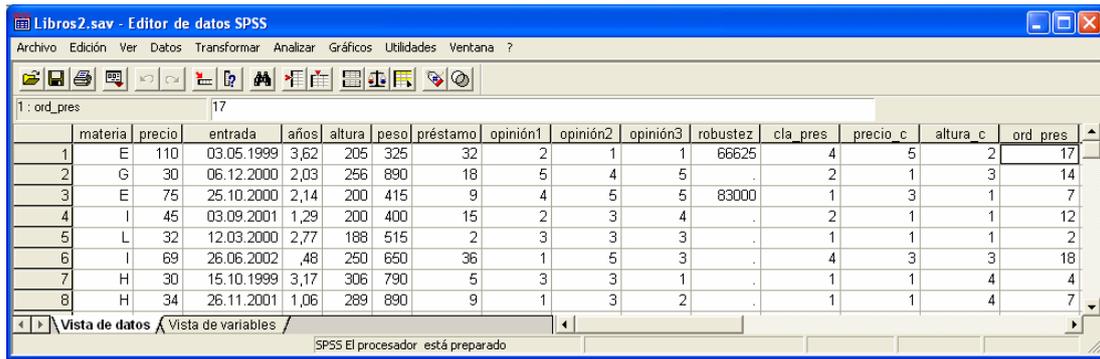


Figura 37: Editor de datos con la nueva variable ord\_pres

En la carpeta **Vista de datos** (Figura 37) podemos observar, por ejemplo, que el resultado de la variable **préstamo** para el caso (libro) número 1 es igual a **32** y que este resultado ocuparía el lugar número **17** si ordenáramos los datos de dicha variable de menor a mayor (éste es el dato del caso número 1 de la nueva variable **ord\_pres**).

Para terminar, podemos volver a guardar el archivo **Libros2.sav**

### 1.13.6. Ponderación de los casos

**Observación** Los datos de este apartado no se grabarán en el archivo **Libros2.sav** sino en un nuevo archivo, cuyo nombre se dará.

La opción **Datos ⇒ Ponderar casos** proporciona a los casos diferentes ponderaciones para el análisis estadístico.

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.

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

Cuadro 3: Datos que, posteriormente, vamos a introducir y grabar con el nombre **PalabrasPorLinea.sav**

Vamos a crear un nuevo archivo 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 Cuadro 3 correspondientes al número de palabras por línea de una página de un libro. El número de palabras por línea es la variable y el número de líneas indica la frecuencia absoluta.

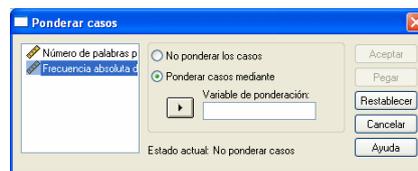
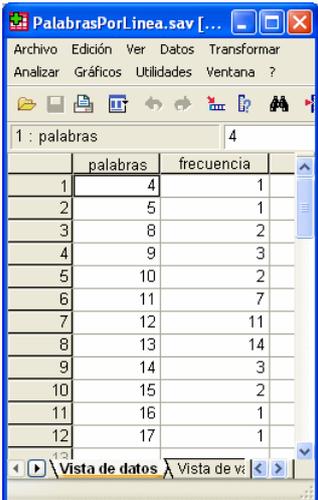


Figura 38: Cuadro de diálogo para ponderar los casos de una variable

Para introducir estos datos hacemos lo siguiente:

- ♥ Seleccionamos la opción **Archivo ⇒ Nuevo ⇒ Datos**.
- ♥ En la carpeta **Vista de variables** definimos la primera variable con las siguientes especificaciones: *Nombre (de variable): palabras; Tipo: Numérica, Anchura= 2, Cifras decimales= 0; Etiqueta (de variable): Número de palabras por línea de texto; Columnas= 8; Medida: Escala*
- ♥ En la misma carpeta definimos la segunda variable (que, en realidad, no es una variable, ya que contiene las frecuencias absolutas de los distintos resultados de la variable **palabras**) con las siguientes especificaciones: *Nombre (de variable): frecuencia; Tipo: Numérica, Anchura= 2, Cifras decimales= 0; Etiqueta (de variable): Frecuencia absoluta de la variable palabras; Columnas= 10; Medida: Escala*
- ♥ En la carpeta **Vista de datos** introducimos los datos de ambas variables a partir del Cuadro 3.



The screenshot shows the SPSS data editor window for the file 'PalabrasPorLinea.sav'. The window title is 'PalabrasPorLinea.sav [...]'. The menu bar includes 'Archivo', 'Edición', 'Ver', 'Datos', 'Transformar', 'Analizar', 'Gráficos', 'Utilidades', and 'Ventana'. The toolbar contains various icons for file operations and data manipulation. The main area displays a data table with two columns: 'palabras' and 'frecuencia'. The table contains 12 rows of data. The status bar at the bottom indicates 'Vista de datos' and 'Vista de v:'. The data is as follows:

	palabras	frecuencia
1	4	1
2	5	1
3	8	2
4	9	3
5	10	2
6	11	7
7	12	11
8	13	14
9	14	3
10	15	2
11	16	1
12	17	1

Figura 39: Ventana del editor de datos con los resultados de la variable palabras

- ♡ Seleccionamos la opción **Datos** ⇒ **Ponderar casos**. Nos aparece el cuadro de diálogo de la Figura 38. En este cuadro de diálogo se activa la opción **Ponderar casos mediante**, y en **Variable de ponderación** se selecciona (mediante la lista que hay a la izquierda y el botón ) la columna **frecuencia**. A continuación se pulsa en **Aceptar**. Así, la apariencia de la carpeta **Vista de datos** será la de la Figura 39.
- ♡ Grabamos este nuevo archivo de datos con el nombre **PalabrasPorLinea.sav**.

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

## 2. Estadística descriptiva unidimensional

### 2.1. Distribución de frecuencias

A partir de este momento, volvemos a trabajar con el archivo **Libros.sav**

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 40. En este cuadro se selecciona, en primer lugar, la variable o variables objeto del procedimiento. Para ello se marcan las



Figura 40: Cuadro de diálogo de la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**

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 *click* sobre los nombres (o las etiquetas) 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 archivo de datos **Libros.sav** seleccionamos la variable **materia** y luego pulsamos el botón **Aceptar** entonces aparece la *ventana del Visor SPSS* de la Figura 41.

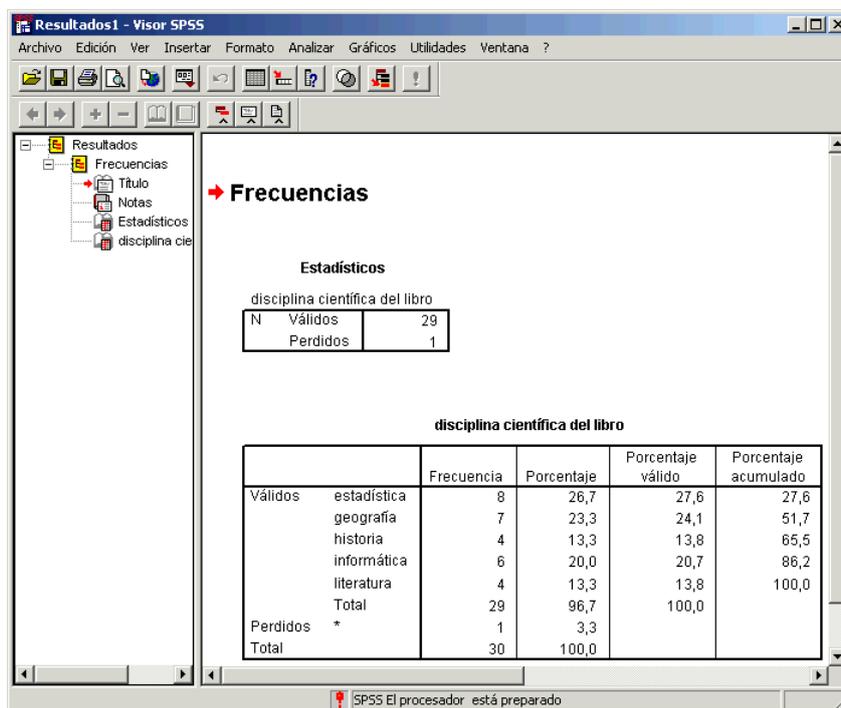


Figura 41: *Visor SPSS* con la distribución de frecuencias de la variable **materia** (*disciplina científica del libro*)

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

### 2.2.1. Cómo guardar y recuperar una ventana del visor SPSS

Antes de cerrar una *ventana del Visor SPSS*, 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 seleccionar **Archivo** ⇒ **Abrir** ⇒ **Resultados**.

### 2.2.2. 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 SPSS* de la Figura 41) 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** ⇒ *Carpeta de Etiquetas de los resultados*. Entonces aparece el cuadro de diálogo de la Figura 42.

Se pueden visualizar los nombres de las variables, las etiquetas de 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 la *ventana del Visor SPSS* 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 42.

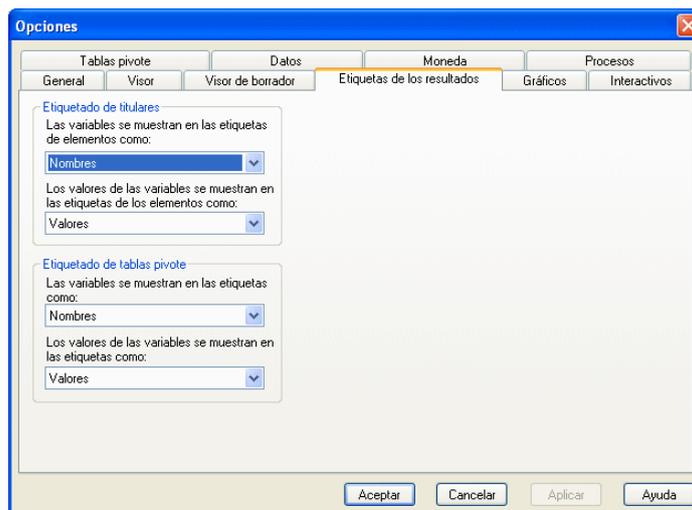


Figura 42: *Carpeta de Etiquetas de los resultados*, dentro de la opción **Edición** ⇒ **Opciones**

Como se ha indicado, los cambios realizados se pondrán de manifiesto en la siguiente *ventana del Visor SPSS* que obtengamos.

### 2.2.3. 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 menús hay que seleccionar **Edición** ⇒ **Opciones** ⇒ *Carpeta de Tablas pivote*. Aparece entonces el cuadro de diálogo de la Figura 43.

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**. Por defecto, la lista muestra los aspectos de tabla guardados en el sub-directorio *Looks* del directorio donde se instaló *SPSS*. 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 en la *ventana del Visor SPSS* (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:

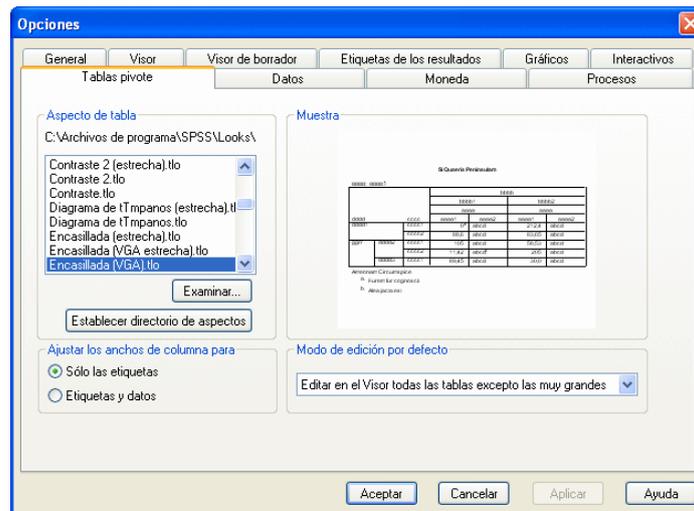


Figura 43: Carpeta de Tablas pivote, dentro de la opción Edición ⇒ Opciones

- **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).
- **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 SPSS* o en una ventana independiente. Por defecto, cuando se pulsa dos veces en una tabla pivote, ésta se activa en la *ventana del Visor SPSS*. 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 SPSS* y las más grandes en una ventana independiente.

Los cambios realizados en el aspecto de las tablas pivote se pondrán de manifiesto la siguiente vez que obtengamos una tabla pivote en la *ventana del Visor SPSS*.

### 2.2.4. Cómo cambiar las especificaciones del visor SPSS por defecto

Para establecer las especificaciones de la *ventana del Visor SPSS* por defecto hay que seleccionar **Edición ⇒ Opciones ⇒ Carpeta de Visor**. Aparece entonces el cuadro de diálogo de la Figura 44. Las opciones de salida en la *ventana del Visor*

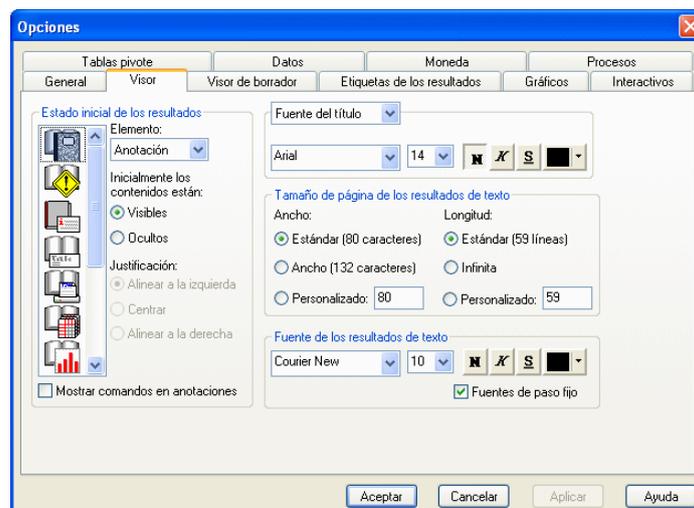


Figura 44: Carpeta de Visor, dentro de la opción Edición ⇒ Opciones

SPSS sólo afectan a los resultados obtenidos tras el cambio de la configuración. A los resultados que ya se muestran en la *ventana del Visor SPSS* 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: anotación, advertencias, notas, títulos, título de página, tablas pivote, gráficos, resultados de texto, mapa, modelo de árbol y árbol de titulares.

Todos los elementos de resultados aparecen alineados a la izquierda en la *ventana del Visor SPSS*. 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.

A la derecha de **Estado inicial de los resultados** hay un recuadro en el cual se puede seleccionar entre **Fuente del título** ó **Fuente del título de página**.

**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.

**Fuente 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.

## 2.3. Medidas descriptivas de los datos

### 2.3.1. Mediante la opción *Analizar* ⇒ *Estadísticos descriptivos* ⇒ *Frecuencias*

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 de la Figura 40, la variable o variables de las cuales se quiere calcular sus medidas descriptivas, y a continuación se pulsa en el botón **Estadísticos**. Por ejemplo, con los datos del archivo **Libros.sav**, en el cuadro de diálogo de la Figura 40 seleccionamos la variable **altura** (haciendo *clic* sobre el nombre o la etiqueta de la variable y *clic* sobre el botón ) , dejamos activada la opción **Mostrar tablas de frecuencias** y pulsamos el botón **Estadísticos**.

Entonces aparece el cuadro de diálogo de la Figura 45. En este cuadro tenemos:

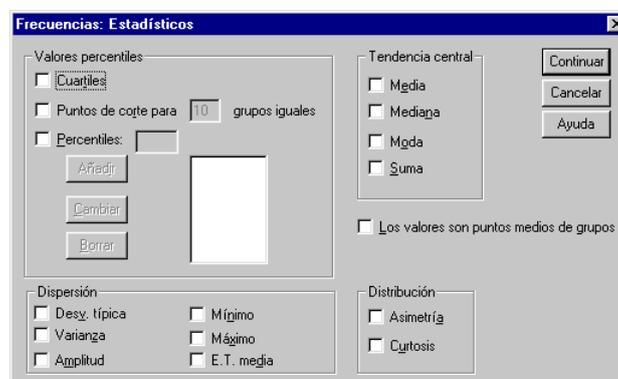


Figura 45: Cuadro de diálogo para seleccionar las medidas descriptivas o estadísticos

**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,  $s^2$ , ya que:

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

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 46: Medidas descriptivas o estadísticos de la variable altura

Continuando con nuestro ejemplo (determinación de las medidas descriptivas de los datos de la variable **altura**), en el cuadro de diálogo de la Figura 45 seleccionamos todas los estadísticos o medidas descriptivas, excepto **Puntos de corte para**  **grupos iguales**, **Percentiles** y **Los valores son puntos medios de grupos**. Pulsamos **Continuar** y luego **Aceptar**. Entonces, en la *ventana del Visor SPSS* aparecen dos tablas: una con las medidas descriptivas o estadísticos (tabla de la Figura 46) y otra con la distribución de frecuencias (tabla de la Figura 47).

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

abriendo el archivo de datos **PalabrasPorLinea.sav**. Recordemos que en dicho archivo también está grabada la ponderación que hicimos en el apartado 1.13.6: los datos de la primera columna (que contiene los resultados de la variable) están ponderados mediante la segunda columna (que tiene 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, una vez abierto el el archivo de datos **PalabrasPorLinea.sav**, se selecciona la opción **Analizar** ⇒ **Estadísticos**

		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 47: Distribución de frecuencias de la variable altura

**descriptivos** ⇒ **Frecuencias** y en el cuadro de diálogo de la Figura 40 marcamos 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 45. En él seleccionamos los estadísticos o medidas descriptivas que deseemos, por ejemplo, los cuartiles, los deciles, 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 archivo 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 45; y el botón **Aceptar** en el de la Figura 40 nos aparece, en la *ventana del Visor SPSS*, la tabla con la distribución de frecuencias (tabla de la Figura 48) y la tabla con los estadísticos seleccionados (tabla de la Figura 49). En este ejemplo, como el tamaño muestral es  $n = 48$  y 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 48: Distribución de frecuencias de la variable palabras

### 2.3.2. Mediante la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Descriptivos**

Las medidas descriptivas se pueden obtener también mediante la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Descriptivos**.

Si volvemos al archivo de datos **Libros.sav** y seleccionamos la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Descriptivos**, nos aparece entonces el cuadro de diálogo de la Figura 50. Por ejemplo, seleccionamos las variables **precio**, **altura** y **peso**

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
	75	13,00
	80	13,00
90	14,10	

Figura 49: Medidas descriptivas o estadísticos de la variable palabras



Figura 50: Cuadro de diálogo de la opción Analizar ⇒ Estadísticos descriptivos ⇒ Descriptivos

(haciendo *clic* sobre cada nombre o etiqueta de cada variable y *clic* sobre el botón ) y pulsamos en **Aceptar**. Entonces se calculan sólo las siguientes medidas descriptivas (de las tres variables elegidas): N (tamaño muestral), Mínimo, Máximo, Media y Desv. típica (cuasidesviación típica). Si quisiéramos obtener otras medidas descriptivas distintas de las anteriores, pulsaríamos el botón **Opciones** del cuadro de diálogo de la Figura 50. Aparecería, entonces, el cuadro de la Figura 51. En



Figura 51: Cuadro de diálogo de las Opciones dadas en Analizar ⇒ Estadísticos descriptivos ⇒ Descriptivos

este cuadro, lo único que es distinto de lo que aparece en el cuadro de la Figura 45 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.

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 puede calcular y guardar los valores tipificados de cada variable seleccionada. Los valores tipificados de la variable  $X$  son los resultados de la transformación  $Z=[X-\text{Media}]/\text{Desviación típica}$ . Cuando las variables se registran en unidades diferentes, una transformación de este tipo pondrá las variables en una escala común para una comparación visual más fácil. Para calcular y guardar los valores tipificados se activa la opción **Guardar valores tipificados como variables** en el cuadro de diálogo de la Figura 50. Cuando se guardan los valores tipificados, éstos se añaden a los datos en la carpeta **Vista de datos** de la *ventana del editor de datos*, quedando disponibles para los gráficos, el listado de los datos y los análisis.

### 2.3.3. Mediante la opción *Analizar* ⇒ *Estadísticos descriptivos* ⇒ *Explorar*

Una tercera forma de obtener las medidas descriptivas es mediante la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Explorar**. Con este procedimiento se generan estadísticos de resumen y representaciones gráficas, ó para todos los casos ó de forma separada para grupos de casos. Existen numerosas razones para utilizar este procedimiento: para inspeccionar los datos, identificar valores atípicos, obtener descripciones, comprobar supuestos y caracterizar diferencias entre subpoblaciones (grupos de casos). La inspección de los datos puede mostrar que existen valores inusuales, valores extremos, discontinuidades en los datos u otras peculiaridades. La exploración de los datos puede ayudar a determinar si son adecuadas las técnicas estadísticas que está teniendo en consideración para el análisis de los datos. La exploración puede indicar que necesita transformar los datos si la técnica necesita una distribución normal. O bien, el usuario puede decidir que necesita utilizar pruebas no paramétricas.

A partir de nuestro archivo **Libros.sav**, vamos a obtener las principales medidas descriptivas o estadísticos de la variable **precio** para cada uno de los cinco grupos (E=estadística, G=geografía, I=informática, L=literatura, H=historia) de la variable **materia**. Para ello, seleccionamos la opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Explorar**. Nos aparece, entonces, el cuadro de diálogo de la Figura 52. En este cuadro de diálogo seleccionamos la variable **precio** para ponerla en **Depen-**



Figura 52: Cuadro de diálogo de la opción *Analizar* ⇒ *Estadísticos descriptivos* ⇒ *Explorar*

**dientes** (haciendo *clic* sobre su nombre o etiqueta y *clic* sobre el botón  que hay delante del recuadro **Dependientes**) y seleccionamos la variable **materia** para ponerla en **Factores** (haciendo *clic* sobre su nombre o etiqueta y *clic* sobre el botón  que hay delante del recuadro **Factores**); activamos la opción **Mostrar ... Estadísticos** y pulsamos el botón **Aceptar**. Entonces obtenemos la tabla de la Figura 53.

La opción **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Explorar** ofrece muchas más posibilidades, que se pueden consultar en la *Ayuda* de *SPSS*.

## 2.4. Representaciones gráficas unidimensionales

**Observación** El aspecto de cualquier representación gráfica puede modificarse (consultar el apartado 2.4.8). De hecho, los gráficos que aparecen en este apartado no son los que aparecen por defecto, sino que han sido modificados.

### 2.4.1. Representaciones gráficas mediante la opción *Analizar* ⇒ *Estadísticos descriptivos* ⇒ *Frecuencias*

Una de las formas de obtener las representaciones gráficas más usuales de las distribuciones de frecuencias es seleccionar **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias**.

Con los datos del archivo **Libros.sav** vamos a obtener el diagrama de barras de frecuencias absolutas de la variable **altura**. Como hemos dicho anteriormente, en primer lugar hay que elegir, en el cuadro de diálogo de la Figura 40, la variable o variables de las cuales se quiere obtener su representación gráfica. En nuestro caso, seleccionamos la variable **altura** para ponerla en el recuadro **Variables** y luego pulsamos el botón **Gráficos**. Entonces aparece el cuadro de diálogo de la Figura 54. En este cuadro tenemos:

MATERIA		Estadístico	Error tip.			
PRECIO	E	Media	71,63	10,200		
		Intervalo de confianza para la media al 95%	Límite inferior	47,51		
			Límite superior	95,74		
		Media recortada al 5%	71,08			
		Mediana	64,00			
		Varianza	832,268			
		Desv. tip.	28,849			
		Mínimo	43			
		Máximo	110			
		Rango	67			
		Amplitud intercuartil	58,00			
		Asimetría	,336	,752		
		Curtosis	-2,118	1,481		
		G	G	Media	53,71	6,421
Intervalo de confianza para la media al 95%	Límite inferior			38,00		
	Límite superior			69,43		
Media recortada al 5%	53,46					
Mediana	57,00					
Varianza	288,571					
Desv. tip.	16,987					
Mínimo	30					
Máximo	82					
Rango	52					
Amplitud intercuartil	24,00					
Asimetría	,243			,794		
Curtosis	,416			1,587		
H	H			Media	42,50	8,098
		Intervalo de confianza para la media al 95%	Límite inferior	16,73		
			Límite superior	68,27		
		Media recortada al 5%	41,89			
		Mediana	37,00			
		Varianza	262,333			
		L	L	Desv. tip.	16,197	
				Mínimo	30	
				Máximo	66	
				Rango	36	
				Amplitud intercuartil	28,50	
				Asimetría	1,631	1,014
				Curtosis	2,709	2,619
				Media	52,67	3,712
Intervalo de confianza para la media al 95%	Límite inferior			43,13		
	Límite superior			62,21		
Media recortada al 5%	52,19					
Mediana	49,50					
Varianza	82,667					
Desv. tip.	9,092					
Mínimo	45					
Máximo	69					
Rango	24					
Amplitud intercuartil	14,25					
Asimetría	1,431	,846				
Curtosis	1,682	1,741				
Media	38,00	2,449				
Intervalo de confianza para la media al 95%	Límite inferior	30,20				
	Límite superior	45,80				
Media recortada al 5%	38,11					
Mediana	39,00					
Varianza	24,000					
Desv. tip.	4,899					
Mínimo	32					
Máximo	42					
Rango	10					
Amplitud intercuartil	9,00					
Asimetría	-,544	1,014				
Curtosis	-2,944	2,619				

Figura 53: Medidas descriptivas de la variable precio para cada uno de los grupos definidos por la variable materia

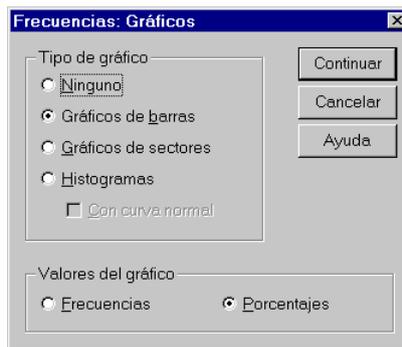


Figura 54: Cuadro de diálogo para seleccionar el tipo de gráfico

**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.

Continuando con nuestro ejemplo (obtención del diagrama de barras de frecuencias absolutas de la variable **altura**) en el apartado **Tipo de gráfico** seleccionamos **Gráficos de barras**; y en el apartado **Valores del gráfico** elegimos **Frecuencias**. Pulsamos **Continuar** y en el cuadro de diálogo de la Figura 40 desactivamos la opción **Mostrar tablas de frecuencias** pues no queremos que aparezca la distribución de frecuencias de la variable **altura**. Por último, pulsamos en **Aceptar**. Nos aparece, entonces, en la *ventana del Visor SPSS*, el gráfico de la Figura 55.

Ahora abrimos los datos del archivo **PalabrasPorLinea.sav**, que contiene los resultados 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

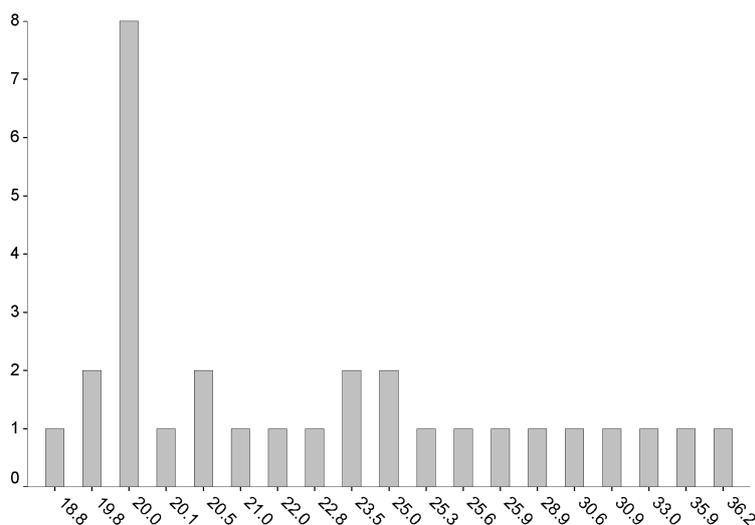


Figura 55: Diagrama de barras de frecuencias absolutas de la variable altura

Vamos a obtener el diagrama de barras de porcentajes de la variable **palabras**. Para ello, una vez abierto el archivo **PalabrasPorLinea.sav**, seleccionamos **Analizar** ⇒ **Estadísticos descriptivos** ⇒ **Frecuencias** y en el cuadro de diálogo de la Figura 40 marcamos la variable **palabras** (o su etiqueta) para ponerla en el recuadro **Variables**. En este cuadro de diálogo pulsamos el botón **Gráficos** y aparece el cuadro de diálogo de la Figura 54. 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 SPSS*, el gráfico de la Figura 56.

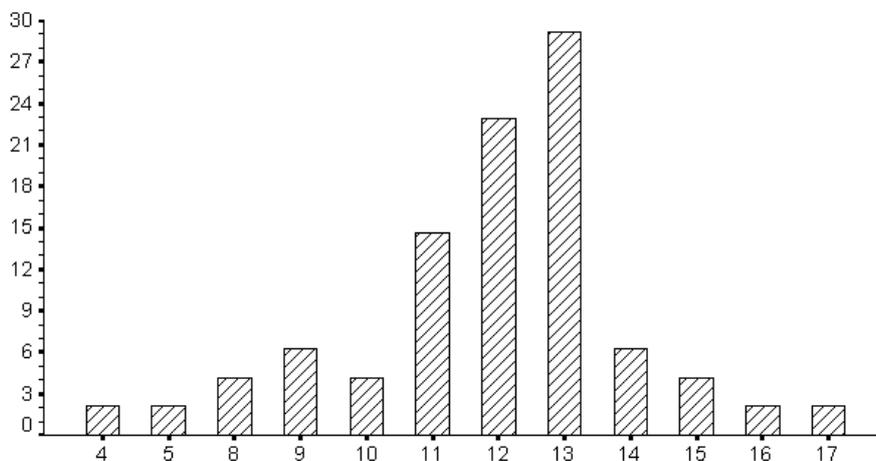


Figura 56: Diagrama de barras de porcentajes de la variable palabras

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

Ahora que hemos hecho nuestros primeros gráficos, quisiera señalar que las especificaciones generales de las representaciones gráficas se pueden cambiar mediante **Edición** ⇒ **Opciones** ⇒ **Carpetas de Gráficos**. Nos aparece entonces el cuadro de diálogo de la Figura 57. Los distintos apartados de este cuadro de diálogo son:

**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

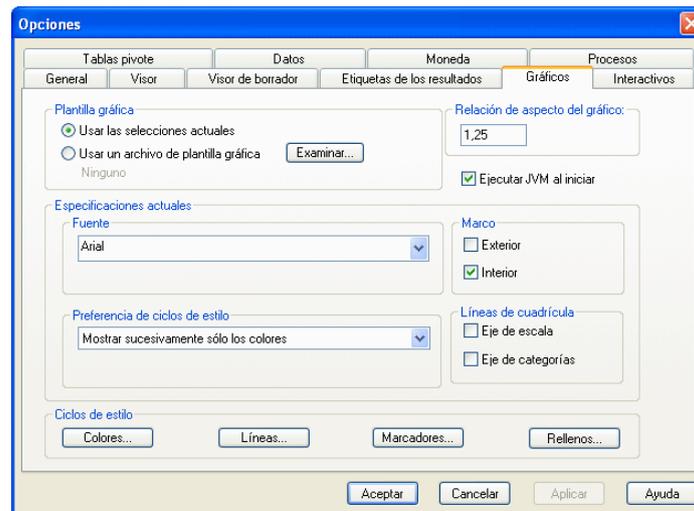


Figura 57: Carpeta de Gráficos, dentro de la opción Edición ⇒ Opciones

en los gráficos, y de 1,8 si las hay.) En este cuadro de diálogo la separación decimal hay que ponerla con una coma abajo.

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

**Preferencia de ciclos de estilo.** Hace la asignación inicial de colores y tramas para nuevos gráficos:

**Mostrar sucesivamente sólo los colores** utiliza solamente los colores para diferenciar los elementos de los gráficos.

**Mostrar sucesivamente sólo las tramas** utiliza solamente las tramas de relleno para diferenciar los elementos de los gráficos.

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

**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.

**Ciclos de estilo.** Personaliza los colores, estilos de línea, símbolos de marcador y tramas de relleno para los gráficos nuevos. Puede cambiar el orden de los colores o tramas utilizados al crear un gráfico nuevo.

### 2.4.3. Creación de diagramas de barras mediante el menú de *Gráficos*

En este apartado vamos a trabajar otra vez con el archivo de datos **Libros.sav**.

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**.

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 hay varias posibilidades. Nosotros vamos a trabajar con la opción **Gráficos** ⇒ **Cuadros de diálogo antiguos** ⇒ **Barras**. Con esta opción aparece el cuadro de diálogo de la Figura 58.

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:

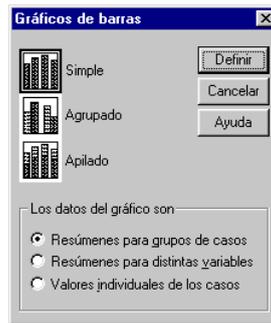


Figura 58: Cuadro de diálogo para hacer un diagrama de barras mediante Gráficos ⇒ Cuadros de diálogo antiguos ⇒ Barras

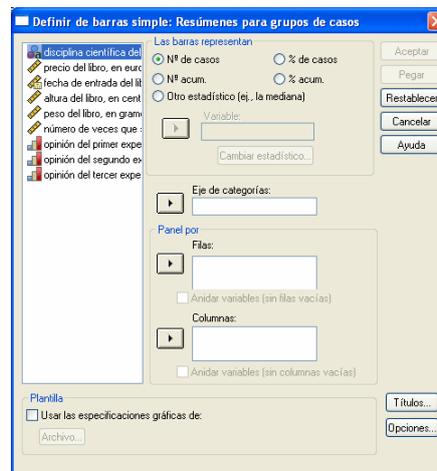


Figura 59: Cuadro de diálogo para hacer un diagrama de barras Simple con la opción Resúmenes para grupos de casos

**Resúmenes para grupos de casos** (del cuadro de diálogo de la Figura 58): Cada barra representa la frecuencia de cada categoría de una variable.

Por ejemplo, con el archivo de datos **Libros.sav** vamos a hacer el diagrama de barras de porcentajes de la variable **materia** (disciplina científica del libro). Para ello, seleccionamos **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Barras**. En el cuadro de diálogo de la Figura 58 seleccionamos el icono **Simple**; en *Los datos del gráfico son* activamos la opción **Resúmenes para grupos de casos** y pulsamos en **Definir**. Entonces aparece el cuadro de diálogo de la Figura 59. En este cuadro de diálogo, en **Eje de categorías** se selecciona la variable cuya distribución de frecuencias se desea representar. Por defecto, haciendo *clic* 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 acum.** Frecuencia absoluta acumulada de cada categoría.

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

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

**Otro estadístico (ej., la mediana).** El eje vertical, así como la altura de cada barra se asocia a otra variable que es especificada en el recuadro **Variable**. Haciendo *clic* en el botón **Cambiar estadístico** se accede al cuadro de diálogo de la Figura 68, que se examinará en el apartado 2.4.4.

Continuando con el ejemplo (determinación del diagrama de barras de porcentajes de la variable **materia**), en **Eje de categorías** del cuadro de diálogo de la Figura 59 ponemos la variable **materia** (*clic* sobre el nombre o la etiqueta de la variable y *clic* sobre el botón  que hay a la izquierda de **Eje de categorías**); en **Las barras representan** seleccionamos **% de casos**; pulsamos en **Opciones** y, en el cuadro de diálogo resultante (Figura 70), seleccionamos **Mostrar los grupos definidos por los valores perdidos**; pulsamos en **Continuar** y después en **Aceptar**. Entonces obtenemos el diagrama de barras de porcentajes correspondiente a la variable **materia**, que es el gráfico de la Figura 60.

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

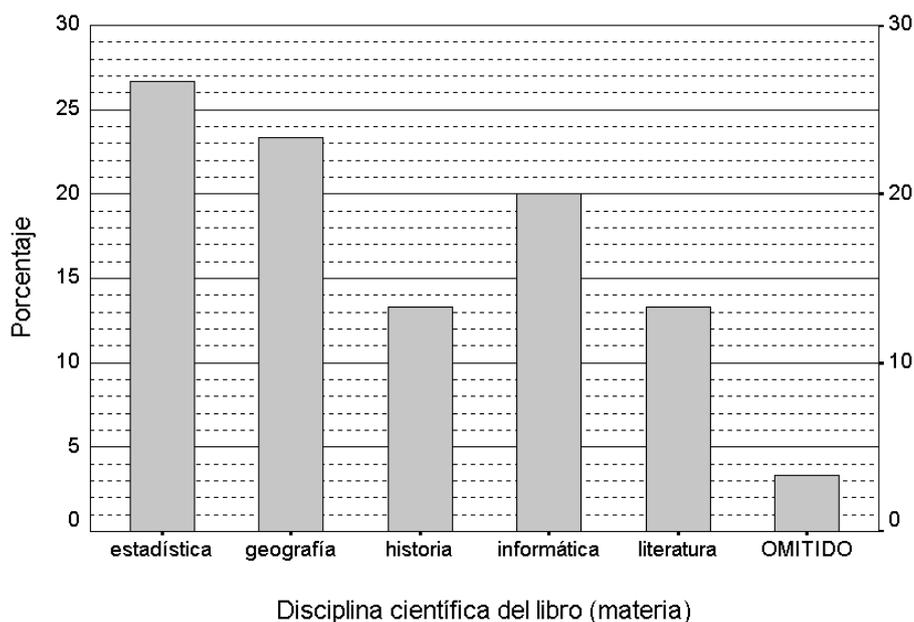


Figura 60: Diagrama de barras de porcentajes de la variable materia

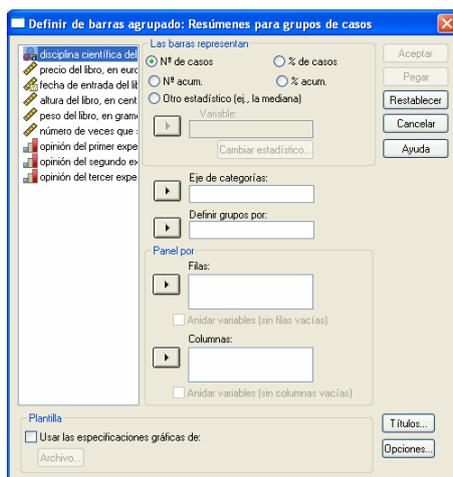


Figura 61: Cuadro de diálogo para hacer un diagrama de barras Agrupado con la opción Resúmenes para grupos de casos

Figura 61); 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.

Por ejemplo, con el archivo de datos **Libros.sav** vamos a a hacer el diagrama de barras de la variable **materia** en grupos definidos por la variable **opinión1**. Para ello, seleccionamos **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Barras**. En el cuadro de diálogo de la Figura 58 seleccionamos el icono **Agrupado**; en *Los datos del gráfico son* activamos la opción **Resúmenes para grupos de casos** y pulsamos en **Definir**. Entonces aparece el cuadro de diálogo de la Figura 61. En **Eje de categorías** de dicho cuadro de diálogo ponemos la variable **materia**; en **Definir grupos por** seleccionamos la variable **opinión1**; pulsamos en **Opciones** y, en el cuadro de diálogo resultante (Figura 70), seleccionamos **Mostrar los grupos definidos por los valores perdidos**; pulsamos en **Continuar** y después en **Aceptar**. Entonces obtenemos el diagrama de barras de frecuencias absolutas de la Figura 62.

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 **opinión1**, que es el gráfico de la Figura 63.

**Resúmenes para distintas variables** (del cuadro de diálogo de la Figura 58): 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.

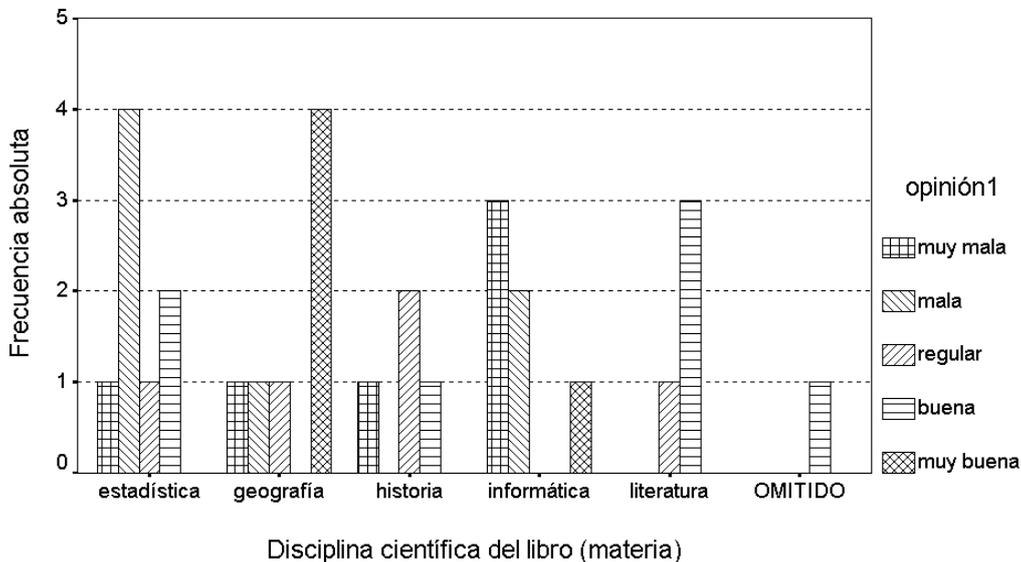


Figura 62: Diagrama de barras de la variable materia en grupos definidos por la variable opinión1

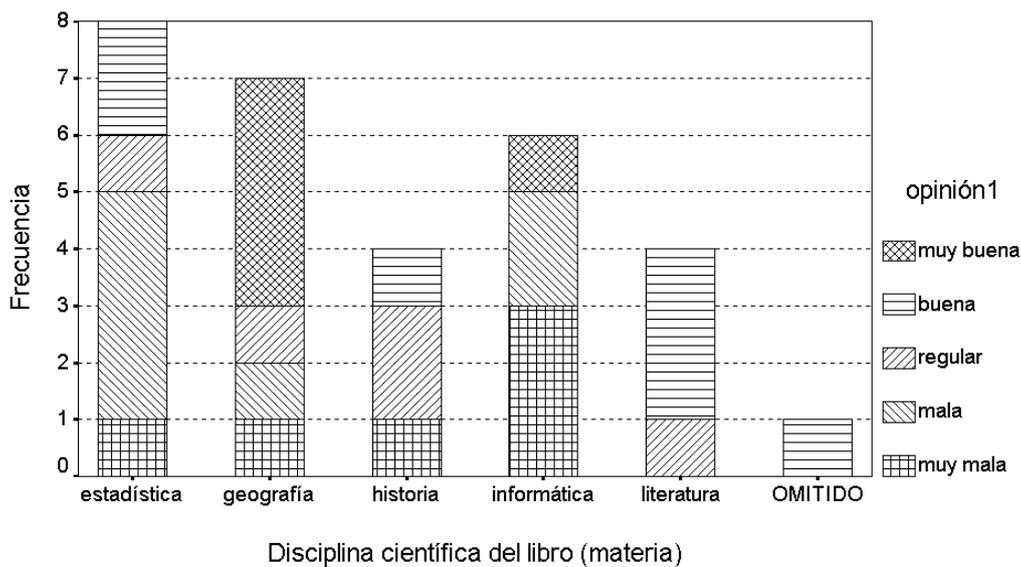


Figura 63: Diagrama de barras de la variable materia apiladas en grupos definidos por la variable opinión1

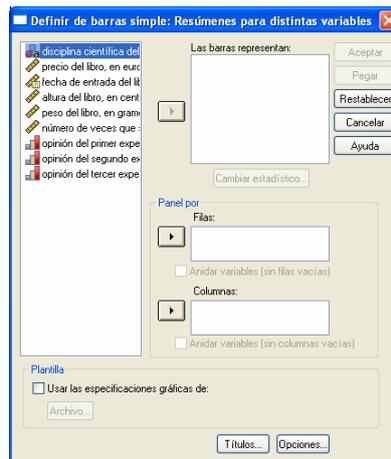


Figura 64: Cuadro de diálogo para hacer un diagrama de barras Simple con la opción Resúmenes para distintas variables

Por ejemplo, con el archivo de datos **Libros.sav** vamos a hacer un diagrama de barras para comparar las medianas de las variables **opinión1**, **opinión2** y **opinión3**. Para ello, seleccionamos **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Barras**. En el cuadro de diálogo de la Figura 58 seleccionamos el icono **Simple**; en *Los datos del gráfico son* activamos la opción **Resúmenes para distintas variables** y pulsamos en **Definir**. Entonces aparece el cuadro de diálogo de la Figura 64. En el recuadro que hay bajo **Las barras representan** de dicho cuadro de diálogo ponemos las tres variables **opinión1**, **opinión2** y **opinión3** (*clic* sobre el nombre o la etiqueta de cada variable y *clic* sobre el botón  que hay a la izquierda de dicho recuadro). Por defecto, cada barra representa la media aritmética de cada variable seleccionada; pero como nosotros queremos comparar las medianas (no las medias) tenemos que cambiar el estadístico que representa cada barra. Para hacerlo, hacemos *clic* sobre **MEAN(opinión del primer...)** y pulsamos en **Cambiar estadístico**. Nos aparece, entonces, el cuadro de diálogo de la Figura 68. En dicho cuadro de diálogo elegimos **Mediana de valores** y pulsamos en **Continuar**. Entonces volvemos al cuadro de diálogo de la Figura 64 y tenemos que hacer lo mismo con la segunda variable; es decir, hacemos *clic* sobre **MEAN(opinión del segundo...)**; pulsamos en **Cambiar estadístico**; elegimos **Mediana de valores** y pulsamos en **Continuar**. Por último, hacemos lo mismo con la tercera variable; es decir, hacemos *clic* sobre **MEAN(opinión del tercer...)**; pulsamos en **Cambiar estadístico**; elegimos **Mediana de valores** y pulsamos en **Continuar**. Tras pulsar en **Aceptar** obtenemos el diagrama de barras para comparar las medianas de las variables **opinión1**, **opinión2** y **opinión3** (Figura 65).

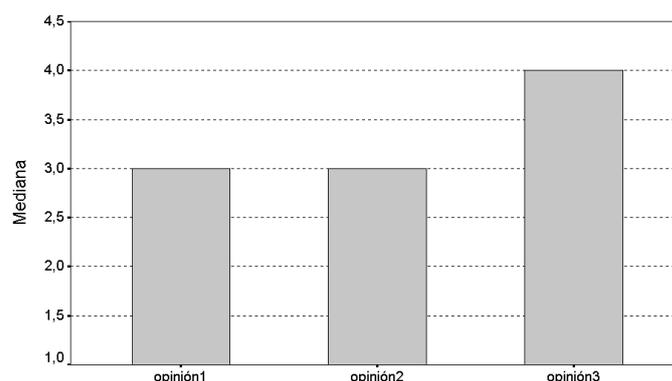


Figura 65: Comparación de las medianas de las variables **opinión1**, **opinión2** y **opinión3**

En los diagramas de barras agrupadas y apiladas dentro de la opción **Resúmenes para distintas variables** 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** (del cuadro de diálogo de la Figura 58): Cada barra representa un caso, lo que quiere decir que en muestras grandes la gráfica sería muy difícil de interpretar.

Por ejemplo, con el archivo de datos **Libros.sav** vamos a hacer un diagrama de barras en el cual las barras repre-



Figura 66: Cuadro de diálogo para hacer un diagrama de barras Simple con la opción Valores individuales de los casos

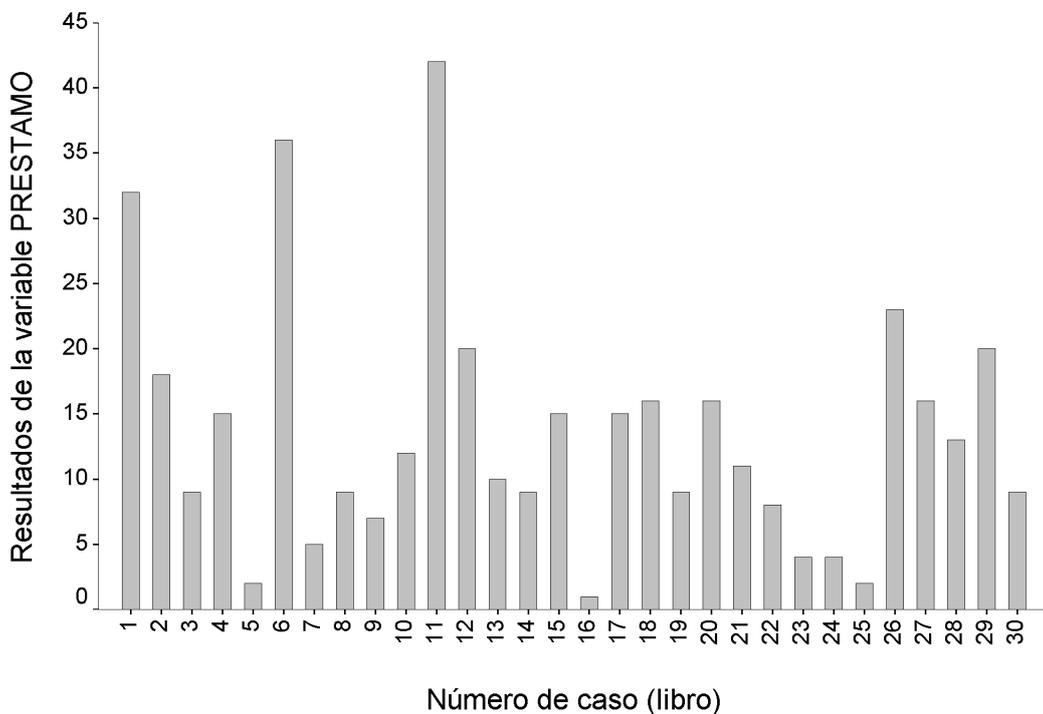


Figura 67: Las barras representan los resultados de la variable préstamo para cada caso (libro)

senten los resultados de la variable **préstamo** para cada caso (libro). Para ello, seleccionamos **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Barras**. En el cuadro de diálogo de la Figura 58 seleccionamos el icono **Simple**; en *Los datos del gráfico son* activamos la opción **Valores individuales de los casos** y pulsamos en **Definir**. Entonces aparece el cuadro de diálogo de la Figura 66. En **Las barras representan** de dicho cuadro de diálogo 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 nuestro ejemplo, seleccionamos la variable **préstamo** (*clic* sobre su nombre o etiqueta y *clic* sobre el botón ). 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 nuestro ejemplo, dejamos activada la opción **Número del caso** y pulsamos en **Aceptar**. De esta forma obtenemos el diagrama de barras de la Figura 67.

En los casos de diagramas agrupados o apilados con la opción **Valores individuales de los caso** hay que seleccionar, evidentemente, más de una variable. En estos casos, en la leyenda adjunta al diagrama se mostrarían los



Figura 68: Cuadro de diálogo para cambiar el estadístico

nombres de las variables.

Se puede observar que en algunos cuadros de diálogo (por ejemplo, el de la Figura 64), se encuentran dos botones, uno referido a **Opciones** y otro referido a **Títulos** que para prácticamente todos los tipos de gráficas son análogos. Estas cuestiones se tratarán en los apartados 2.4.5 y 2.4.6.

#### 2.4.4. Cuadro de diálogo de *Cambiar estadístico*

Como se ha dicho, haciendo *click* en **Cambiar estadístico** situado en los cuadros de diálogo de algunos tipos de gráficas (por ejemplo, los de las Figuras 59, 61 y 64) se accede al cuadro de diálogo de la Figura 68. Aquí se presentan las siguientes alternativas, referidas a la variable o variables declaradas en el recuadro en que **Cambiar estadístico** 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.

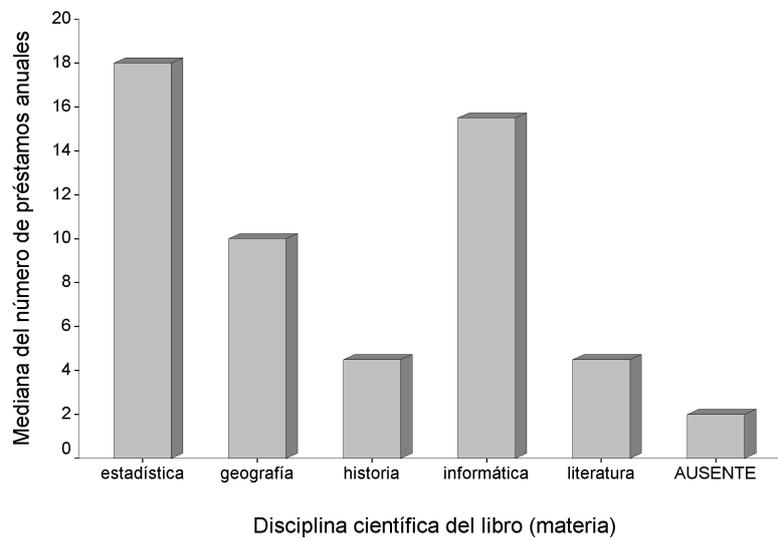


Figura 69: Diagrama de barras de la variable materia respecto de la Mediana de la variable préstamo

Por ejemplo, con el archivo de datos **Libros.sav** vamos a hacer un diagrama de barras de la variable **materia** respecto del estadístico **Mediana** de la variable **préstamo**; es decir, vamos a hacer un gráfico que nos compare las medianas de la variable **número de préstamos anuales** para cada uno de los cinco grupos (E=estadística, G=geografía, I=informática, L=literatura, H=historia) de la variable **materia**. Para ello, seleccionamos **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Barras**. En el cuadro de diálogo de la Figura 58 seleccionamos el icono **Simple**; en *Los datos del gráfico son* activamos la opción **Resúmenes para grupos de casos** y pulsamos en **Definir**. En el cuadro de diálogo que aparece (Figura 59) seleccionamos la variable **materia** (disciplina científica del libro) en **Eje de categorías**. En dicho cuadro activamos **Otro estadístico** y ponemos la variable **préstamo** (número de veces que se ha prestado el libro en el último año) en **Variable**. Pulsamos el botón **Cambiar estadístico** y en el cuadro que aparece (Figura 68) seleccionamos **Mediana de valores**. Por último, pulsamos en **Continuar** y en **Aceptar**. El diagrama de barras resultante es el de la Figura 69.

#### 2.4.5. Cuadro de diálogo de Opciones

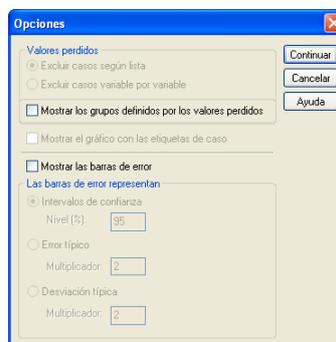


Figura 70: Cuadro de diálogo para seleccionar las Opciones de las representaciones gráficas

Seleccionando el botón de **Opciones** del cuadro de diálogo inicial de algunos tipos de representaciones gráficas (por ejemplo, el cuadro de diálogo de la Figura 64) nos aparece el cuadro de la Figura 70. Las opciones de gráfico disponibles dependen del tipo de gráfico y de los datos.

**Valores perdidos:** sólo se encuentran disponibles cuando un nuevo gráfico va a mostrar o resumir varias variables (no se incluyen las variables que definen grupos).

- **Excluir casos según lista:** excluye los casos del gráfico completamente si incluyen valores perdidos para alguna de las variables resumidas.
- **Excluir casos variable por variable:** excluye por separado los casos de los distintos estadísticos de resumen calculados. Los distintos elementos del gráfico deben estar basados en grupos de casos diferentes.

**Mostrar los grupos definidos por los valores perdidos:** sólo se encuentra disponible cuando se utiliza una variable categórica para definir los grupos de un nuevo gráfico. Si se selecciona, en el gráfico se mostrarán en grupos diferentes los distintos valores perdidos de la variable de agrupación (incluido el valor perdido del sistema). Si no se selecciona, se excluirán del gráfico los casos con valores perdidos del sistema o definidos como perdidos por el usuario para la variable de subgrupo.

**Mostrar el gráfico con las etiquetas de caso:** sólo se encuentra disponible cuando se define un diagrama de dispersión y se ha especificado una variable para etiquetar los casos mediante en el cuadro de diálogo principal. Si se selecciona, se mostrarán en el gráfico como etiquetas de punto las etiquetas de valor (o los valores en caso de que no se hayan definido etiquetas) de la variable especificada.

#### 2.4.6. Cuadro de diálogo de *Títulos*



Figura 71: Cuadro de diálogo para colocar *Títulos* a las representaciones gráficas

Haciendo *clic* en el botón **Títulos** del cuadro de diálogo inicial de cada tipo de gráfica (cuadros de diálogo de las Figuras 59, 61, 64 y 66) se muestra un cuadro como el de la Figura 71. 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.

#### 2.4.7. Creación de histogramas mediante el menú de *Gráficos*

Una de las formas de realizar un histograma es mediante la opción **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Histograma**. Entonces aparece el cuadro de diálogo de la Figura 72. 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**, con la opción **Mostrar curva normal** seleccionada, es el de la Figura 73.



Figura 72: Cuadro de diálogo para hacer un histograma

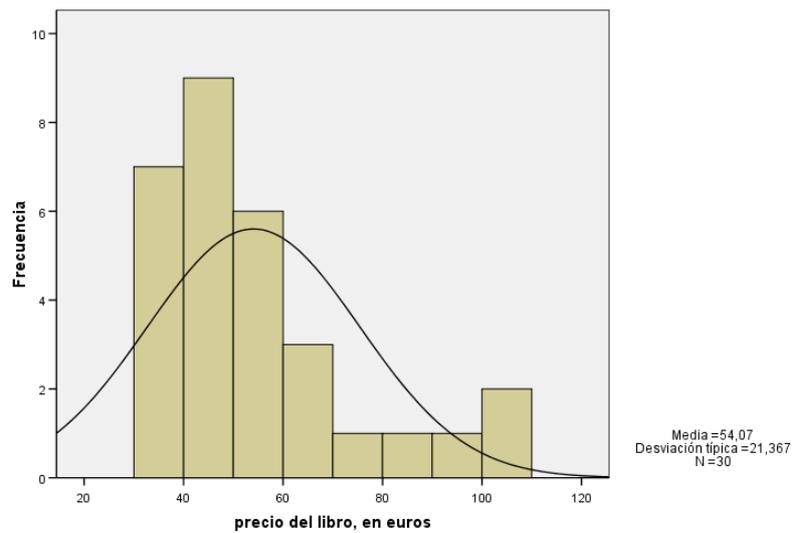


Figura 73: Histograma correspondiente a la variable precio (sin modificaciones)

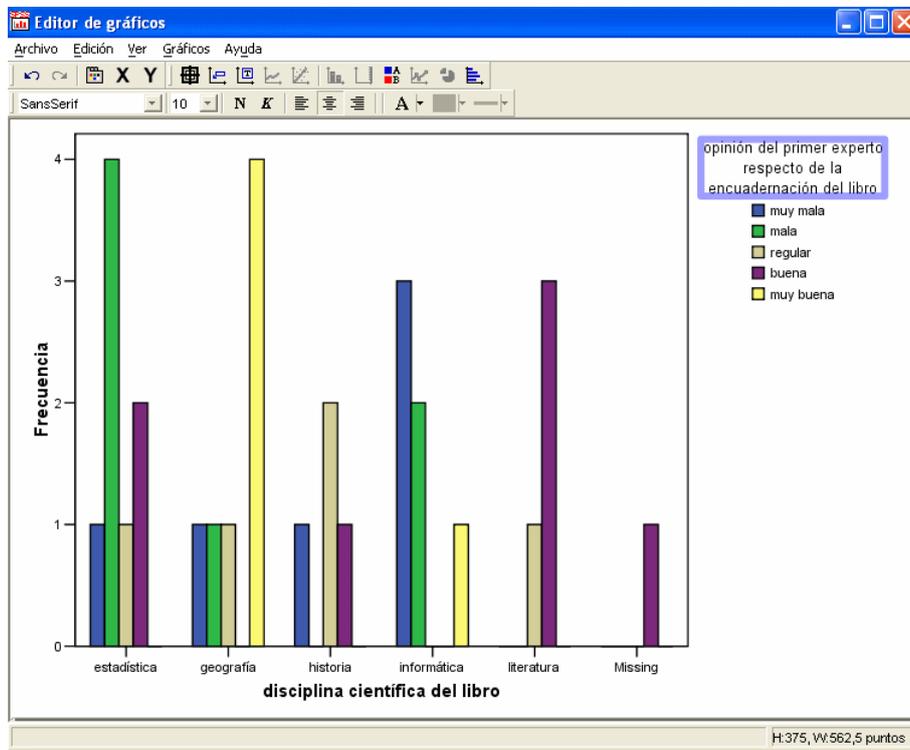


Figura 74: Ventana del editor de gráficos con un diagrama de barras agrupadas

### 2.4.8. Modificación de gráficos

Cualquier gráfico puede editarse y modificarse haciendo doble *clic* dentro de la zona donde está representado en la *ventana del Visor SPSS*. Al hacerlo, entramos en la ventana del *editor de gráficos*.

El *editor de gráficos* proporciona un entorno potente y fácil de usar en el que se puede personalizar los gráficos y explorar los datos.

Vamos a volver a hacer el diagrama de barras de la variable **materia** en grupos definidos por la variable **opinión1** del archivo de datos **Libros.sav** (gráfico de la Figura 62). El gráfico que sale por defecto (sin modificaciones) es el de la Figura 74. Hacemos doble *clic* dentro de la zona donde está representado el gráfico en la *ventana del Visor SPSS*. Al hacerlo, entramos en la ventana del *editor de gráficos*.

Una vez abierta la ventana del *editor de gráficos*, se puede modificar el gráfico con los menús, con las barras de herramientas o haciendo doble *clic* en el objeto que se desea modificar. Para seleccionar objetos en el gráfico, se hace *clic*

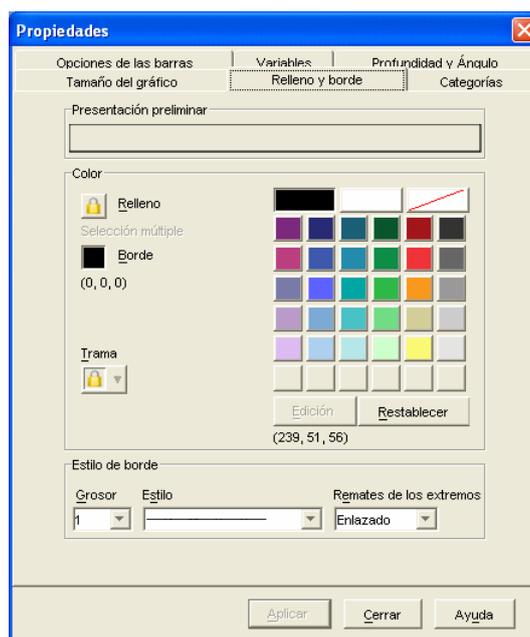


Figura 75: Ventana para cambiar las propiedades de un diagrama de barras

sobre ellos. Por ejemplo, si se hace *clik* sobre el título de la leyenda, aparecerá un *recuadro de selección* alrededor de él (ver otra vez la Figura 74).

Para modificar un gráfico utilizando los menús de la ventana del *editor de gráficos*, lo más habitual es seleccionar la parte del gráfico a editar y elegir **Edición** ⇒ **Propiedades**. En nuestro ejemplo, seleccionamos todas las barras de la Figura 74 haciendo *clik* sobre una de ellas, y elegimos **Edición** ⇒ **Propiedades**. Se obtiene el cuadro de diálogo de la Figura 75 (cuando está abierta la pestaña o carpeta **Relleno y borde**).

La ventana **Propiedades** de la Figura 75 incluye una serie de pestañas que permiten definir y realizar los distintos cambios en el gráfico. Las pestañas que se pueden ver en dicha ventana se basan en la selección actual realizada en el gráfico. Algunas pestañas incluyen una presentación preliminar para que se tenga una idea del aspecto de los cambios. No obstante, el gráfico no reflejará los cambios hasta que se pulse en **Aplicar**. Se pueden realizar cambios en más de una pestaña antes de pulsar **Aplicar**. En función de la selección, es posible que no todas las propiedades estén disponibles. La ayuda para cada una de las pestañas especifica qué se debe seleccionar para ver las pestañas y cambiar las propiedades concretas de dichas pestañas. Si se han seleccionado varios elementos (por ejemplo, todas las barras del diagrama de la Figura 74) sólo se pueden cambiar las propiedades comunes a todos ellos.

La pestaña **Tamaño del gráfico** de la ventana **Propiedades** de la Figura 75 permite introducir las dimensiones exactas del marco exterior. La pestaña **Relleno y borde** permite cambiar los colores, aplicar una trama de relleno o modificar el ancho de línea o estilo del borde del marco exterior, marco de los datos, marco de la leyenda, cuadros de texto y barras. La pestaña **Categorías** permite reordenar y excluir categorías en un gráfico de sectores o un gráfico con un eje de categorías. La pestaña **Profundidad y ángulo** permite añadir efectos de tres dimensiones a los gráficos de barras. La pestaña **Variables** permite *cambiar el papel* de las variables en diversos tipos de gráficos. El *cambio de papel* de las variables incluye el desplazamiento de una variable del eje independiente al eje dependiente y viceversa y cambio de la variable de agrupación. La introducción de estos cambios altera el modo en que el *editor de gráficos* calcula los estadísticos. No se trata de un simple cambio de aspecto del gráfico. La pestaña **Opciones de las barras** es específica de este tipo de gráfico, y permite cambiar el ancho de los elementos de datos de barra y las agrupaciones de barras.

Las opciones **Seleccionar eje X**, **eje Y** o **eje Z** del menú **Edición** de la ventana del *editor de gráficos* nos llevan a un cuadro de diálogo de **Propiedades** similar al de la Figura 76. Las pestañas de esta ventana de **Propiedades** nos permiten gobernar la **Marcas y cuadrículas** del gráfico, así como las **Etiquetas del eje** seleccionado. Con la pestaña **Categorías** se pueden reordenar o excluir categorías en un gráfico de sectores o un gráfico con un eje categórico. La pestaña **Tamaño del gráfico** es similar a la comentada anteriormente.

Otra forma de modificar un gráfico es usar las *barras de herramientas* de la ventana del *editor de gráficos*. Estas barras de herramientas proporcionan métodos abreviados para algunas de las configuraciones de la ventana de *Propiedades*.

Por otra parte, también es posible modificar una representación gráfica haciendo doble *clik* sobre el elemento que se quiere editar. Por ejemplo, para cambiar las propiedades de la etiqueta **Frecuencia** de uno de los ejes del gráfico de la Figura 74 se hace doble *clik* sobre dicha etiqueta y se accede, entonces, a un cuadro de diálogo como el de la Figura 77.

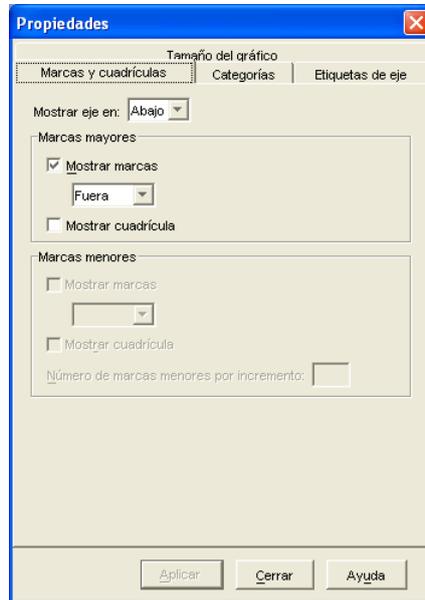


Figura 76: Ventana para cambiar las propiedades de uno de los ejes de un gráfico



Figura 77: Ventana para cambiar las propiedades de la etiqueta de uno de los ejes

Para cambiar, por ejemplo, las marcas señalizadoras de los ejes se hace doble *clic* sobre una de ellas y, entonces, se accede a un cuadro de diálogo similar al de la Figura 76.

A fin de practicar con el *editor de gráficos*, se pueden modificar los gráficos que aparecen por defecto para llegar a la forma en que aparecen en las secciones 2.4.1, 2.4.3 y 2.4.4 (gráficos de las Figuras 55, 56, 60, 62, 63, 65, 67 y 69).

### 3. Relación entre dos variables

#### 3.1. Relación entre variables cuantitativas. Diagrama de dispersión

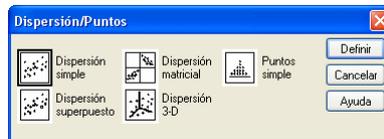


Figura 78: Cuadro de diálogo para elegir el tipo de diagrama de dispersión

Para obtener el diagrama de dispersión o nube de puntos se selecciona la opción **Gráficos**⇒**Cuadros de diálogo antiguos**⇒**Dispersión/Puntos** en la que se presentan cinco alternativas (ver la Figura 78): *Dispersión simple*, *Dispersión superpuesta*, *Dispersión matricial*, *Dispersión 3-D* y *Puntos simple*. De las cinco, la que hemos estudiado nosotros es la primera de ellas.

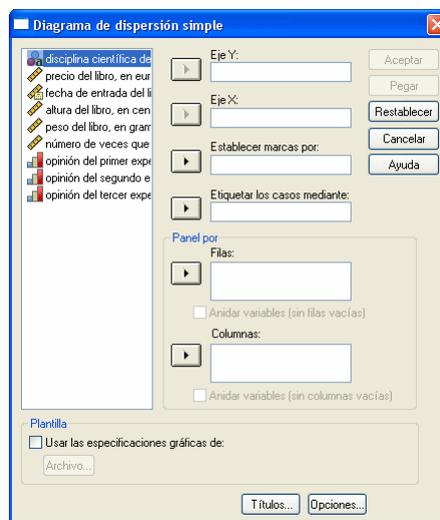


Figura 79: Cuadro de diálogo para especificar las características de un diagrama de dispersión simple

Por tanto, en el cuadro de la Figura 78 se selecciona el icono **Dispersión simple** y se pulsa en **Definir**. Se presenta un nuevo cuadro de diálogo (Figura 79) en el cual se especifican las dos variables: una correspondiente al eje vertical (**Eje Y**) y la otra al eje horizontal (**Eje X**).

Por ejemplo, del archivo de datos **Libros.sav** vamos a hacer el diagrama de dispersión de la variable **precio** sobre la variable **peso**. Para ello, en el cuadro de diálogo de la Figura 79 ponemos la variable **precio** en el recuadro de **Eje Y** (*clic* sobre su nombre o etiqueta y *clic* sobre el botón que hay a la izquierda de **Eje Y**), ponemos la variable **peso** en el recuadro de **Eje X** (*clic* sobre su nombre o etiqueta y *clic* sobre el botón que hay a la izquierda de **Eje X**) y pulsamos en **Aceptar**. Obtenemos, entonces, el gráfico de la Figura 80.

En el recuadro **Establecer marcas por** del cuadro de la Figura 79 se puede especificar una nueva variable, habitualmente cualitativa, para que cada punto se vea representado por un símbolo distinto según el valor o categoría que adopte en la mencionada variable. Por ejemplo, el diagrama de dispersión de la variable **precio** sobre la variable **peso**, seleccionando la variable **materia** en **Establecer marcas por** es el de la Figura 81.

En el recuadro **Etiquetar los casos mediante** del cuadro de la Figura 79 se puede seleccionar otra nueva variable para adjuntar a cada punto del diagrama el valor que le corresponde en la variable ahora declarada. Esta opción hace que el gráfico parezca muy confuso, por lo que es conveniente no utilizarla.

Mediante **Opciones** del cuadro de diálogo de la Figura 79 podemos especificar el tratamiento de los valores perdidos en los datos y determinar si se muestran etiquetas para los puntos del diagrama (ver el apartado 2.4.5).

Con la opción **Títulos** del cuadro de diálogo de la Figura 79 se pueden definir las líneas de texto que se colocan en la parte superior o inferior del diagrama (ver el apartado 2.4.6).

#### 3.2. Modificación del diagrama de dispersión

Una vez obtenido el diagrama de dispersión y para proceder a su modificación hay que editarlo haciendo doble *clic* dentro de la representación gráfica en la *ventana del Visor SPSS*. Al hacerlo, entramos en la ventana del *editor de gráficos*.

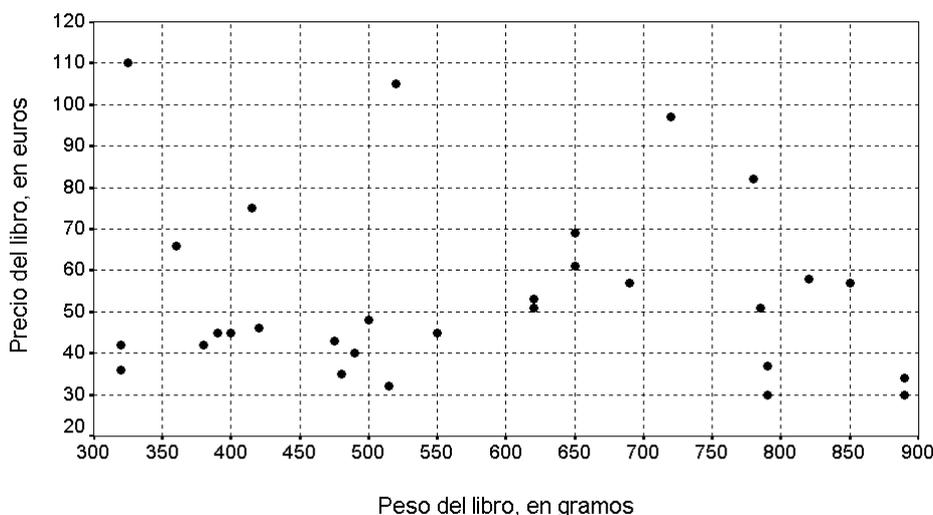


Figura 80: Diagrama de dispersión de la variable precio sobre la variable peso

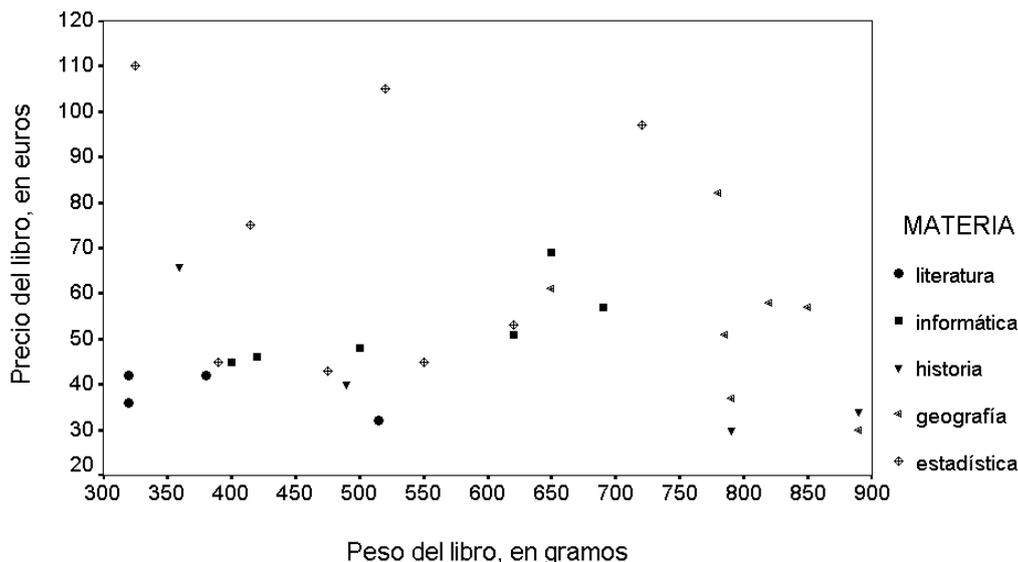


Figura 81: Diagrama de dispersión del precio sobre el peso con marcas distintas según la variable materia

Las posibilidades de modificación del diagrama de dispersión son similares a las del diagrama de barras (ver apartado 2.4.8), por lo que voy a exponer sólo algunas de las particularidades de la edición del diagrama de dispersión.

Una posibilidad de cambio importante es la modificación de los símbolos que aparecen en dicho gráfico. En *SPSS*, a estos símbolos se les denomina *marcadores*. Por defecto, en esta versión de *SPSS*, los marcadores son circunferencias muy pequeñas. Se puede cambiar tanto la forma del marcador de datos como su tamaño, color y ancho de borde. Estos cambios se pueden hacer, por ejemplo, de la siguiente manera: se hace doble *clic* dentro de la representación gráfica en la *ventana del Visor SPSS* para entrar en el *editor de gráficos* y se hace doble *clic* sobre uno de los marcadores que se quieren modificar. Con ello se produce una doble función; por un lado, se seleccionan todos los marcadores del mismo tipo, y por otro lado, se abre la ventana de **Propiedades** (Figura 82). En la pestaña de **Marcador** de esta ventana se pueden distinguir los siguientes apartados:

**Presentación preliminar.** El área de presentación preliminar muestra las opciones actuales y el aspecto que presentará el marcador o marcadores una vez que se haya pulsado en **Aplicar**.

**Marcador.** Se cambia el tipo de marcador, su tamaño (en puntos) o el grosor de su borde (en puntos). Por regla general, se debe evitar los bordes gruesos a menos que el tamaño del marcador también sea grande. Utilizar bordes gruesos

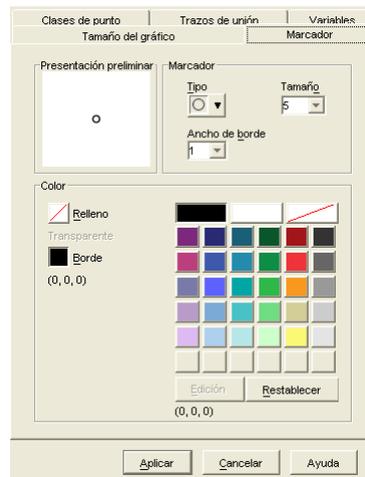


Figura 82: Ventana para modificar los marcadores de un diagrama de dispersión

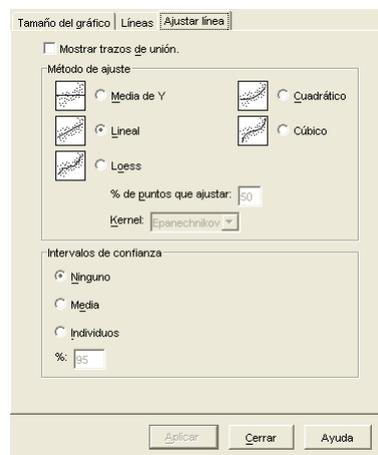


Figura 83: Ventana para añadir una línea de ajuste en un diagrama de dispersión

en marcadores pequeños puede ocultar la forma del marcador y el color del relleno.

**Color.** Se selecciona un color o una trama para el relleno y/o para el borde. Para ello, se pulsa en el *cuadrado* situado junto al relleno o borde, y después se selecciona, a la derecha, el color o la trama deseada. Los cambios del marcador de color y las configuraciones de rojo, verde y azul (R, V, A) aparecen debajo del marcador. Hay que tener en cuenta que el color blanco con una línea roja diagonal que lo atraviesa equivale a transparente. Para personalizar un color, se selecciona dicho color y se pulsa en **Edición** para abrir un cuadro de diálogo estándar donde personalizar los colores. Se puede utilizar la última fila para crear colores personalizados. Para restablecer todos los colores a su configuración original se pulsa en **Restablecer**.

Otra modificación importante en el diagrama de dispersión es la posibilidad de incluir una línea de ajuste (por ejemplo, la recta de regresión).

Para añadir una línea de ajuste, en la ventana del *editor de gráficos* se seleccionan todos los marcadores haciendo *clic* sobre uno de ellos. A continuación, en la barra de menús se selecciona **Elementos** ⇒ **Línea de ajuste total**. Aparece, entonces, la ventana de **Propiedades** de la Figura 83. En esta ventana se utiliza la pestaña **Ajustar línea** para especificar las opciones de la línea de ajuste. Las opciones son:

**Mostrar trazos de unión.** Dibuja trazos de unión de cada punto de los datos al punto correspondiente de la línea de ajuste.

**Media de y.** Dibuja una recta horizontal situada en la media de los valores de los datos del eje Y.

**Lineal.** Dibuja la **recta de regresión** determinada mediante el método de mínimos cuadrados. Ésta es la opción que nosotros utilizaremos habitualmente.

**Cuadrático.** Dibuja la curva de regresión procedente de una ecuación segundo grado, determinada mediante el método de mínimos cuadrados.

**Cúbico.** Dibuja la curva de regresión procedente de una ecuación tercer grado, determinada mediante el método de mínimos cuadrados.

**Loess.** Dibuja una línea de ajuste mediante mínimos cuadrados ponderados iterativos. Son necesarios al menos 13 puntos de datos. Este método ajusta un porcentaje especificado de los puntos de los datos siendo el valor predeterminado un 50 %. Además de cambiar el porcentaje, puede seleccionar una función de *kernel* concreta. El *kernel* por defecto (función de probabilidad) funciona correctamente con la mayoría de los datos.

**Intervalos de confianza.** Dibuja un par de líneas alrededor de la línea de ajuste de regresión para ilustrar el nivel porcentual de confianza especificado. La opción **Media** dibuja los intervalos basados en la media pronosticada a partir de la línea de regresión. La opción **Individuos** dibuja los intervalos basados en los valores Y individuales pronosticados a partir de la línea de regresión. Los intervalos de confianza no están disponibles para las líneas de ajuste *Loess* debido a que no se basan en la regresión.

En la ventana de **Propiedades** de la Figura 83 se puede utilizar la pestaña **Líneas** para especificar el formato de la línea de ajuste. Al acabar se pulsa en **Aplicar**. Es posible modificar las propiedades de la línea de ajuste una vez que se se ha añadido en el diagrama de dispersión. Para ello, se selecciona dicha línea de ajuste (haciendo *click* sobre ella) y, en la barra de menús del *editor de gráficos*, se elige **Edición** ⇒ **Propiedades**. Aparece, otra vez, la ventana de la Figura 83.

Por ejemplo, el diagrama de dispersión de la variable **precio** sobre la variable **peso**, con la recta de regresión visualizada, es el gráfico de la Figura 84.

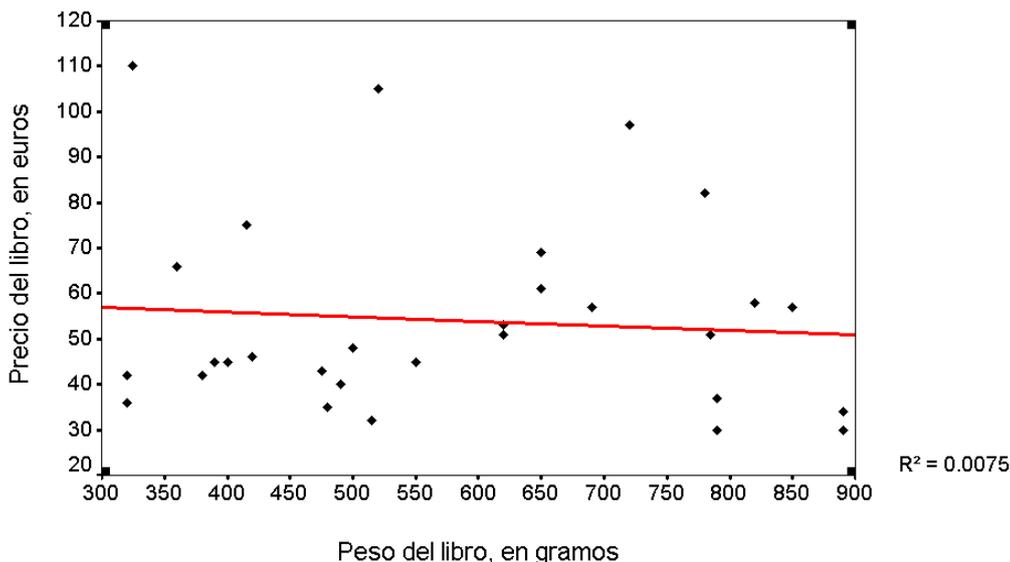


Figura 84: Diagrama de dispersión y recta de regresión de la variable precio sobre la variable peso

### 3.3. Correlación simple



Figura 85: Cuadro de diálogo obtener coeficientes de correlación

Para obtener el coeficiente de correlación lineal de Pearson (y otros coeficientes de correlación) se selecciona **Analizar** ⇒ **Correlaciones** ⇒ **Bivariadas**. Aparece, entonces, el cuadro de diálogo de la Figura 85. En el recuadro de la izquierda de este cuadro de diálogo está la lista de variables, de las cuales podemos seleccionar dos o más. Según sea la naturaleza de las variables, podemos obtener otros coeficientes de correlación en el grupo **Coefficientes de correlación**:

**Pearson:** Calcula el coeficiente de correlación lineal de Pearson entre cada par de variables cuantitativas.

**Tau-b de Kendall:** Calcula el coeficiente de correlación de Kendall entre cada par de variables ordinales (es similar, en su interpretación, al coeficiente de correlación de Spearman).

**Spearman:** Calcula el coeficiente de correlación de Spearman entre cada par de variables ordinales.

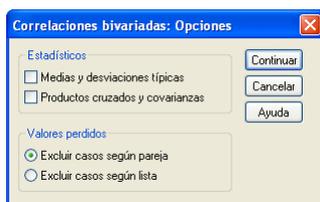


Figura 86: Opciones de Correlaciones Bivariadas

En el cuadro de diálogo de la Figura 85 también aparece un apartado, denominado **Prueba de significación**, que se entenderá cuando se haya estudiado la parte de *Estadística Inferencial*, pues hace un test de hipótesis para contrastar la hipótesis nula de que el coeficiente de correlación, en toda la población, sea igual a cero. En relación con esto, la opción **Marcar las correlaciones significativas** coloca unos asteriscos junto al coeficiente que no resulte igual a cero en la población al aplicarle el contraste de hipótesis (existe correlación entre ese par de variables cuando se considera toda la población).

Por ejemplo, del archivo **Libros.sav** vamos a calcular el coeficiente de correlación lineal de Pearson entre cada par de variables del trío de variables **peso**, **precio** y **altura**. Para ello, seleccionamos **Analizar** ⇒ **Correlaciones** ⇒ **Bivariadas**. En el cuadro de diálogo resultante (Figura 85) seleccionamos las tres variables mencionadas haciendo *click* sobre sus nombres (o sus etiquetas) y *click* sobre el botón para ponerlas en el recuadro que hay debajo de **Variables** y activamos la opción **Pearson**.

Estando situados en este cuadro de diálogo (Figura 85), pulsando en el botón **Opciones** nos encontramos con dos bloques (Figura 86); el primero se refiere a **Estadísticos**:

**Medias y desviaciones típicas:** Calcula las medias aritméticas ( $\bar{x}$ ,  $\bar{y}$ ) y las cuasidesviaciones típicas ( $S_x$ ,  $S_y$ ). A estas últimas, el sistema les llama desviaciones típicas. También se muestra el número de casos que no tienen valores perdidos. Los valores perdidos se consideran según cada variable individual, sin tener en cuenta la opción elegida para la manipulación de los valores perdidos.

**Productos cruzados y covarianzas:** Se calcula lo que el sistema llama *Suma de cuadrados y productos cruzados*, que no es más que la suma de los productos cruzados de las variables corregidas respecto a la media; es decir,

$$\sum_{i=1}^r \sum_{j=1}^k (x_i - \bar{x})(y_j - \bar{y}) f_{ij}.$$

Éste es el numerador de la covarianza.

También se calcula lo que el sistema llama covarianza, pero que en realidad es la cuasicovarianza (similar a la covarianza, pero dividiendo por  $(n - 1)$  en vez de por  $n$ ; siendo  $n$  el tamaño muestral). La cuasicovarianza,  $S_{xy}$ , está relacionada con la covarianza,  $s_{xy}$ , de la siguiente manera:

$$S_{xy} = \frac{\sum_{i=1}^r \sum_{j=1}^k (x_i - \bar{x})(y_j - \bar{y}) f_{ij}}{n - 1} = \frac{n}{n - 1} s_{xy}.$$

De esto se deduce que el coeficiente de correlación lineal de Pearson se puede calcular de cualquiera de las dos formas:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{S_{xy}}{S_x S_y}.$$

En el segundo bloque (del cuadro de diálogo de la Figura 86) se realiza el tratamiento de los valores ausentes, **Valores perdidos**:

**Excluir casos según pareja:** Por defecto, se excluyen los casos con valores ausentes en alguna (o las dos) de las variables cuyo coeficiente de correlación se trata de calcular.

**Excluir casos según lista:** Se excluyen los casos con valores ausentes en alguna de las variables declaradas en la lista de variables del cuadro de diálogo inicial.

	Media	Desviación típica	N
PESO	580,17	184,935	30
PRECIO	54,07	21,367	30
ALTURA	23.820	5.0531	30

Figura 87: Medias y cuasidesviaciones típicas

		PESO	PRECIO	ALTURA
PESO	Correlación de Pearson	1	-.086	,729**
	Sig. (bilateral)	.	,650	,000
	Suma de cuadrados y productos cruzados	991824,167	-9905,333	19767,400
	Covarianza	34200,833	-341,563	681,634
	N	30	30	30
PRECIO	Correlación de Pearson	-.086	1	-.134
	Sig. (bilateral)	,650	.	,482
	Suma de cuadrados y productos cruzados	-9905,333	13239,867	-418,040
	Covarianza	-341,563	466,547	-14,415
	N	30	30	30
ALTURA	Correlación de Pearson	,729**	-.134	1
	Sig. (bilateral)	,000	,482	.
	Suma de cuadrados y productos cruzados	19767,400	-418,040	740,488
	Covarianza	681,634	-14,415	25,534
	N	30	30	30

\*\* . La correlación es significativa al nivel 0,01 (bilateral).

Figura 88: Coeficiente de correlación lineal de Pearson (entre otras medidas)

Continuando con nuestro ejemplo (cálculo del coeficiente de correlación lineal de Pearson entre cada par de variables del trío de variables **peso**, **precio** y **altura**), activamos las opciones **Medias y desviaciones típicas** y **Productos cruzados y covarianzas** del cuadro de diálogo de la Figura 86; luego pulsamos en **Continuar** y en **Aceptar**. Como resultado, en la *ventana del Visor SPSS* aparecen las tablas de las Figuras 87 y 88. En la primera tabla se puede ver (entre otras cosas) que la media aritmética de la variable **precio** es 54'07 euros y su cuasidesviación típica es  $S = 21'367$  euros. Por tanto, su desviación típica es:

$$s = \sqrt{\frac{n-1}{n}} S = \sqrt{\frac{29}{30}} 21'367 = 21'008 .$$

En la segunda tabla de la *ventana del Visor SPSS* (Figura 88) se pueden ver los resultados del coeficiente de correlación lineal de Pearson para cada par de estas variables, así como la suma de los productos cruzados y la cuacovarianza. Se puede observar que el coeficiente de correlación lineal de Pearson entre el **peso** y el **precio** de los libros de la muestra es igual a  $-0'086$ . Como este coeficiente está muy próximo a 0, podríamos decir que la fuerza de la relación lineal entre el **peso** y el **precio** (en la muestra) es muy débil. El coeficiente de correlación lineal de Pearson entre el **peso** y la **altura** de los libros de la muestra es igual a  $0'729$ . Como este coeficiente está bastante próximo a 1, podríamos decir que la fuerza de la relación lineal entre el **peso** y la **altura** (en la muestra) es bastante fuerte. Por último, el coeficiente de correlación lineal de Pearson entre el **precio** y la **altura** de los libros de la muestra es igual a  $-0'134$ . Como este coeficiente está muy próximo a cero, podríamos decir que la fuerza de la relación lineal entre el **precio** y la **altura** (en la muestra) es muy débil.



Figura 89: Cuadro de diálogo para realizar la regresión lineal

### 3.4. Regresión lineal

Para obtener la ecuación de la recta de regresión lineal mínimo cuadrática de una variable cuantitativa  $Y$  sobre otra variable cuantitativa  $X$ , se selecciona la opción **Analizar** ⇒ **Regresión** ⇒ **Lineal**. Aparece, entonces, el cuadro de diálogo de la Figura 89.

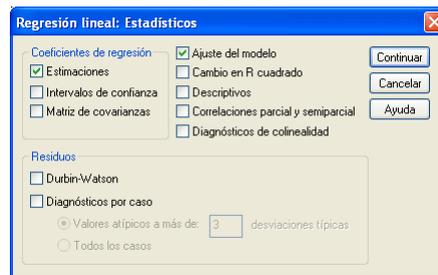


Figura 90: Cuadro de diálogo para obtener, entre otras cosas, los coeficientes de regresión

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	12,257	2,148		5,707	,000
	peso	,020	,004	,729	5,642	,000

a. Variable dependiente: altura

Figura 91: Coeficientes de regresión de la recta de regresión de la variable altura sobre la variable peso

Puesto que hemos obtenido anteriormente que la fuerza de la relación lineal entre el **peso** y la **altura** es bastante fuerte, vamos ahora a encontrar la ecuación de la recta de regresión mínimo cuadrática de la variable **altura** sobre la variable **peso** (de nuestro archivo de datos **Libros.sav**). Para ello, en el cuadro de diálogo de la Figura 89 seleccionamos la variable **altura** en **Dependiente** y la variable **peso** en **Independientes**. Pulsamos el botón **Estadísticos**, en el cuadro de diálogo que aparece (Figura 90), dejamos activada solamente la opción **Estimaciones** del grupo **Coeficientes de regresión** y pulsamos **Continuar**. Volvemos al anterior cuadro de diálogo (Figura 89) y pulsamos **Aceptar**. Así, en la *ventana del Visor SPSS* tenemos la tabla de la Figura 91, en la cual se hace un contraste de hipótesis (que se explicará en la parte de *Estadística Inferencial*), pero lo que a nosotros nos interesa en este momento son los resultados de los coeficientes de regresión, que son:  $A = 12'257$ ,  $B = 0'020$ , siendo la ecuación de la recta de regresión mínimo cuadrática  $Y = A + B X$  (donde  $Y = \text{altura}$  y  $X = \text{peso}$ ). Es decir, la ecuación de dicha recta de regresión es:

$$\text{altura} = 12'257 + 0'020 \cdot \text{peso}$$

Si, además, queremos obtener los valores pronosticados de la variable dependiente a partir de la recta de regresión mínimo cuadrática de la variable dependiente sobre la variable independiente, en el cuadro de diálogo de la Figura 89 pulsamos el botón **Guardar** y, en el cuadro de diálogo que aparece (Figura 92), activamos **Valores pronosticados... No tipificados**. Tras pulsar **Continuar** y **Aceptar**, nos aparece, en la *ventana del editor de datos* una nueva columna (variable) que contiene

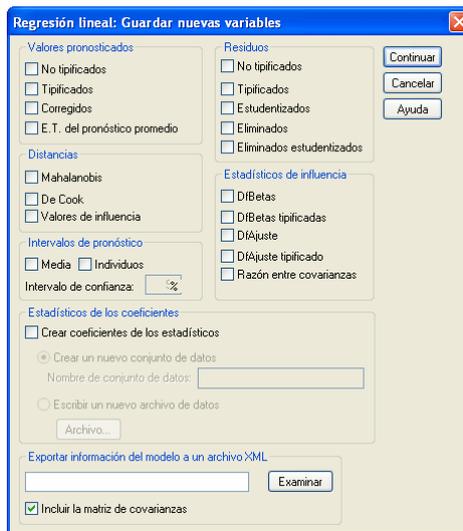


Figura 92: Cuadro de diálogo para guardar algunas medidas, como los valores pronosticados y los residuos o errores

	materia	precio	entrada	altura	peso	préstamo	opinión1	opinión2	opinión3	pre_1	res_1
1	E	110	03.05.1999	20.5	325	32	2	1	1	18,73444	1,76566
2	G	30	06.12.2000	25.6	890	18	5	4	5	29,99609	-4,39609
3	E	75	25.10.2000	20.0	415	9	4	5	5	20,52817	-,52817
4	I	45	03.09.2001	20.0	400	15	2	3	4	20,22922	-,22922
5	L	32	12.03.2000	18.8	515	2	3	3	3	22,52121	-3,72121
6	I	69	26.06.2002	25.0	650	36	1	5	3	25,21180	-,21180
7	H	30	15.10.1999	30.6	790	5	3	3	1	28,00205	2,59795
8	H	34	26.11.2001	28.9	890	9	1	3	2	29,99609	-1,09609
9	L	42	07.07.2000	20.1	320	7	4	2	1	18,63479	1,46521
10	I	46	04.05.2002	20.0	420	12	2	5	3	20,62782	-,62782

Figura 93: Pronósticos de la variable altura a partir de la recta de regresión de dicha variable sobre la variable peso

dichos valores. También se pueden guardar, entre otras cosas, los residuos o errores (las diferencias entre los valores pronosticados y los valores reales de la variable dependiente). Para ello, en el cuadro de diálogo de la Figura 89 pulsamos el botón **Guardar** y, en el cuadro de diálogo que aparece (Figura 92), activamos **Residuos... No tipificados**. Tras pulsar **Continuar** y **Aceptar**, nos aparece, en la *ventana del editor de datos* una nueva columna con dichos residuos o errores en la predicción.

En el ejemplo anteriormente resuelto, los valores pronosticados de la variable **altura** a partir de la recta de regresión de la variable **altura** sobre la variable **peso** aparecen en la penúltima columna (**pre\_1**) de la Figura 93 y los valores de los residuos o errores cometidos en la predicción (diferencia entre **altura** y **pre\_1**) aparecen en la última columna (variable **res\_1**). Como vemos, **SPSS** asigna unos nombres (por defecto) a estas nuevas variables (**pre\_1** y **res\_1**). Es conveniente cambiar estos nombres y poner unas etiquetas explicativas.

## 4. Probabilidad. Variables aleatorias

Para obtener resultados de probabilidad con *SPSS* se tiene que utilizar la opción **Transformar**  $\Rightarrow$  **Calcular variable**, que se ha explicado en la sección 1.13 (sección que es conveniente repasar en este momento).

Para poder usar la opción **Transformar**  $\Rightarrow$  **Calcular variable** es necesario tener abierto un archivo de datos (archivo **.sav**). Podemos utilizar cualquier archivo de datos, pero lo más conveniente es crear uno nuevo con una única variable que tenga un único dato. Para ello, seleccionamos la opción **Archivo**  $\Rightarrow$  **Nuevo**  $\Rightarrow$  **Datos**. Definimos la primera columna dándole sólo el nombre; que puede ser, por ejemplo, **necesaria**. 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). Podemos grabar este archivo de datos con el nombre **Probabilidad.sav**.

Recordemos que si seleccionamos **Transformar**  $\Rightarrow$  **Calcular variable** nos aparece el cuadro de diálogo de la Figura 17. Recordemos también que en este cuadro de diálogo 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 a su derecha está el grupo de funciones disponibles (**Grupo de funciones**). Una vez seleccionado un grupo dentro de la lista **Grupo de funciones**, debajo aparece otra lista con las funciones y variables que se pueden usar dentro del grupo de funciones elegido (**Funciones y variables especiales**).

### 4.1. Función de probabilidad y función de densidad

*SPSS* puede calcular el resultado de la función de probabilidad (si la variable es discreta) o de la función de densidad (si la variable es continua) para un valor concreto.

Como los resultados de la función de probabilidad y de la función de densidad de 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 una *Anchura* igual a 8 y con un número de *Cifras decimales* igual a 6. Para ello, seleccionamos **Edición**  $\Rightarrow$  **Opciones**  $\Rightarrow$  **Carpeta de Datos** de la barra de menús. En el apartado **Visualización para nuevas variables numéricas** del cuadro de diálogo correspondiente ponemos un 8 en **Ancho** y un 6 en **Cifras decimales** y pulsamos en **Aceptar**.

Por ejemplo, sea  $X$  una variable aleatoria Binomial de parámetros  $n = 80$  y  $p = 0'5$ ,  $X \equiv B(80, 0'5)$ . Vamos a calcular el resultado de la función de probabilidad para el valor 30; es decir,  $P(X = 30)$ . Para ello, teniendo abierto el archivo de datos **Probabilidad.sav**, seleccionamos la opción **Transformar**  $\Rightarrow$  **Calcular variable**. En **Variable de destino** podemos escribir **ejemplo1**. Ahora tenemos que poner en el recuadro **Expresión numérica** la función **PDF.BINOM(c,n,p)** que calcula el resultado de la función de probabilidad evaluado en **c** de una variable aleatoria Binomial de parámetros **n** y **p**. Para ello, en **Grupo de funciones** hacemos *click* sobre **FDP y FDP no centrada**; en la lista de **Funciones y variables específicas** hacemos doble *click* sobre **Pdf.Binom**. Podemos observar que en el recuadro que hay a la izquierda de la lista **Funciones y variables específicas** aparece una explicación de lo que hace concretamente la función seleccionada. En **Expresión numérica** aparece lo siguiente: **PDF.BINOM(?, ?, ?)**. Tenemos que sustituir las interrogaciones por lo que queremos; es decir, debe aparecer **PDF.BINOM(30,80,0.5)**. Después de pulsar el botón **Aceptar** obtendremos en la carpeta de datos, a continuación de nuestra última variable, los resultados de la probabilidad pedida,  $P(X = 30)$ , que es igual a 0'007338. Podemos volver a grabar estos nuevos datos en **Probabilidad.sav**.

Entre las funciones de probabilidad y funciones de densidad que aparecen dentro de la opción **FDP y FDP no centrada** del recuadro **Grupo de funciones**, las que hemos estudiado son:

**PDF.BINOM(cant,n,prob)** Calcula el resultado de la función de probabilidad evaluado en **cant** de una variable aleatoria Binomial de parámetros **n** (número de veces que se repite el experimento) y **prob** (probabilidad de éxito).

**PDF.POISSON(cant,media)** Calcula el resultado de la función de probabilidad evaluado en **cant** de una variable aleatoria de Poisson con parámetro **media**.

**PDF.GEOM(cant,prob)** Calcula el resultado de la función de probabilidad evaluado en **cant** de una variable aleatoria geométrica (número de intentos para obtener un éxito) cuando la probabilidad de éxito es **prob**.

**PDF.NORMAL(cant,media,desv\_típ)** Calcula el resultado de la función de densidad evaluado en **cant** de una variable aleatoria normal de parámetros **media** y **desv\_típ**.

**PDF.CHISQ(cant,gl)** Calcula el resultado de la función de densidad evaluado en **cant** de una variable aleatoria chi-cuadrado con los grados de libertad **gl**.

**PDF.T(cant,gl)** Calcula el resultado de la función de densidad evaluado en **cant** de una variable aleatoria t de Student con los grados de libertad **gl**.

**PDF.F(cant,gl1,gl2)** Calcula el resultado de la función de densidad evaluado en **cant** de una variable aleatoria F de Snedecor con los grados de libertad **gl1** y **gl2**.

### Ejercicios Propuestos

1. Sea  $X \equiv B(n = 50, p = 0'25)$ . Calcular  $P(X = 10)$ .
2. Sea  $X \equiv \mathcal{P}(\lambda = 8)$ . Calcular  $P(X = 7)$ .
3. Sea  $X \equiv G(p = 0'25)$ . Calcular  $P(X = 5)$ .

## 4.2. Función de distribución (probabilidad acumulada)

Para calcular el resultado de la función de distribución de una variable aleatoria,  $F(t) = P(X \leq t)$ , también hay que elegir la opción **Transformar**  $\Rightarrow$  **Calcular variable**.

Por ejemplo, vamos a calcular la probabilidad  $P(X \leq -1'36)$ , siendo  $X$  una variable aleatoria Normal Estándar,  $X \equiv \mathcal{N}(0, 1)$ . Como sabemos,  $P(X \leq -1'36) = F(-1'36)$ . Para calcular el resultado de la función de distribución evaluado en  $-1'36$  de una variable aleatoria Normal de media **0** y desviación típica **1**, teniendo abierto el archivo de datos **Probabilidad.sav**, seleccionamos la opción **Transformar**  $\Rightarrow$  **Calcular variable**. En **Variable de destino** podemos escribir **ejemplo2**. Ahora tenemos que poner en el recuadro **Expresión numérica** la función **CDF.NORMAL(cant,media,desv\_típ)** que calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria Normal de parámetros **media** y **desv\_típ**. Para ello, en **Grupo de funciones** hacemos *click* sobre **FDA y FDA no centrada**; en la lista de **Funciones y variables específicas** hacemos doble *click* sobre **Cdf.Normal**. En **Expresión numérica** aparece lo siguiente: **CDF.NORMAL(?,?,?)**. Tenemos que sustituir las interrogaciones por lo que queremos; es decir, debe aparecer **CDF.NORMAL(-1.36,0,1)**. Después de pulsar el botón **Aceptar** obtendremos en la carpeta de datos, a continuación de nuestra última variable, los resultados de la probabilidad pedida,  $P(X \leq -1'36) = F(-1'36) = 0'086915$ . Podemos volver a grabar estos nuevos datos en **Probabilidad.sav**.

Si queremos calcular probabilidades de los tipos  $P(X > a)$ ,  $P(X \geq a)$ ,  $P(a < X < b)$ ,  $\dots$ , tenemos que utilizar lápiz y papel, y aplicar las propiedades de la probabilidad para llegar a expresiones en las que sólo aparezcan probabilidades del tipo  $P(X \leq x)$  (función de distribución), pues éstas son las que calcula **SPSS**. No tenemos que olvidar, por ejemplo, que si  $X$  es una variable aleatoria continua, entonces  $P(X = a) = 0$  para todo  $a$ , por lo que se cumplen las siguientes igualdades:  $P(X \leq a) = P(X < a)$ ,  $P(X \geq a) = P(X > a)$ ,  $\dots$ . Pero si  $X$  es una variable aleatoria discreta, las probabilidades  $P(X \leq a)$  y  $P(X < a)$  no son (en general) iguales.

Por ejemplo, si  $X \equiv \mathcal{N}(6'5, 1'85)$ , entonces  $P(5 \leq X < 7) = P[(X < 7) - (X < 5)] = P(X < 7) - P(X < 5) = P(X \leq 7) - P(X \leq 5) = F(7) - F(5) = CDF.NORMAL(7, 6.5, 1.85) - CDF.NORMAL(5, 6.5, 1.85) = 0'397787$ .

Entre las funciones de distribución que aparecen dentro de la opción **FDA y FDA no centrada** del recuadro **Grupo de funciones**, las que hemos estudiado son:

**CDF.BINOM(cant,n,prob)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria Binomial de parámetros **n** (número de veces que se repite el experimento) y **prob** (probabilidad de éxito).

**CDF.POISSON(cant,media)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria de Poisson con parámetro **media**.

**CDF.GEOM(cant,prob)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria geométrica (número de intentos para obtener un éxito) cuando la probabilidad de éxito es **prob**.

**CDF.NORMAL(cant,media,desv\_típ)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria normal de parámetros **media** y **desv\_típ**.

**CDF.CHISQ(cant,gl)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria chi-cuadrado con los grados de libertad **gl**.

**CDF.T(cant,gl)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria t de Student con los grados de libertad **gl**.

**CDF.F(cant,gl1,gl2)** Calcula el resultado de la función de distribución evaluado en **cant** de una variable aleatoria F de Snedecor con los grados de libertad **gl1** y **gl2**.

### Ejercicios Propuestos

4. Sea  $X \equiv B(n = 50, p = 0'25)$ . Calcular:

- a)  $P(X \leq 12)$ .
- b)  $P(X \geq 3)$ .
- c)  $P(X < 5)$ .
- d)  $P(X > 7)$ .
- e)  $P(10 < X < 20)$ .
- f)  $P(10 \leq X < 20)$ .
- g)  $P(10 < X \leq 20)$ .
- h)  $P(10 \leq X \leq 20)$ .

5. Sea  $X \equiv \mathcal{P}(\lambda = 8)$ . Calcular:

- a)  $P(X < 6)$ .
- b)  $P(X > 7)$ .
- c)  $P(X \leq 5)$ .
- d)  $P(X \geq 9)$ .
- e)  $P(5 < X < 15)$ .
- f)  $P(5 \leq X \leq 15)$ .

6. Sea  $X \equiv G(p = 0'25)$ . Calcular:

- a)  $P(X < 3)$ .
- b)  $P(X > 4)$ .
- c)  $P(X \leq 3)$ .
- d)  $P(X \geq 4)$ .
- e)  $P(3 \leq X < 10)$ .

7. Sea  $X$  una variable aleatoria normal de media 5 y desviación típica 1'5,  $X \equiv \mathcal{N}(5, 1'5)$ . Calcular:

- a)  $P(X \leq 6'5)$ .
- b)  $P(X > 7'5)$ .
- c)  $P(5'5 < X < 7'5)$ .

8. Sea  $X$  una variable aleatoria que sigue una distribución  $t$  de Student con  $n$  grados de libertad,  $X \equiv t_n$ . Calcular:

- a) Para  $n = 10$ ,  $P(X \leq 1'45)$ .
- b) Para  $n = 15$ ,  $P(X \geq 2'5)$ .
- c) Para  $n = 7$ ,  $P(0'57 < X < 3'5)$ .

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

- a) Para  $n = 12$ ,  $P(X < 4'8)$ .
- b) Para  $n = 20$ ,  $P(X > 4'8)$ .
- c) Para  $n = 4$ ,  $P(3'3 < X < 9'4)$ .

10. Sea  $X$  una variable aleatoria con distribución  $F$  de Snedecor con  $m$  y  $n$  grados de libertad,  $X \equiv F_{m,n}$ . Calcular:

- a) Para  $m = 7, n = 5$ ,  $P(X < 2'9)$ .
- b) Para  $m = 25, n = 50$ ,  $P(X \geq 5)$ .
- c) Para  $m = 15, n = 10$ ,  $P(0'4 < X < 2'4)$ .

### 4.3. Inversa de la función de distribución (percentiles o cuantiles)

En ocasiones, en lugar de querer calcular probabilidades de sucesos, se desea justamente lo contrario, conocer el valor  $x$  que hace que la probabilidad del suceso ( $X \leq x$ ) sea igual a un valor determinado  $p$ ; es decir, hallar  $x$  para que se cumpla  $P(X \leq x) = p$ ; esto no es más que calcular percentiles de variables aleatorias.

Como los resultados de los percentiles pueden ser cualesquiera (no tienen que estar 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 una *Anchura* igual a 8 y con un número de *Decimales* igual a 4. Para ello, seleccionamos **Edición**  $\Rightarrow$  **Opciones**  $\Rightarrow$  **Carpeta de Datos** de la barra de menús. En el apartado **Visualización para nuevas variables numéricas** del cuadro de diálogo correspondiente ponemos un 8 en **Ancho** y un 4 en **Cifras decimales**.

Por ejemplo, vamos a calcular el valor  $k$  que verifica  $P(X \leq k) = 0'98$ , cuando  $X$  es una variable aleatoria chi-cuadrado de Pearson con 20 grados de libertad,  $X \equiv \chi_{20}^2$ . Para ello, teniendo abierto el archivo de datos **Probabilidad.sav**, seleccionamos la opción **Transformar**  $\Rightarrow$  **Calcular variable**. En **Variable de destino** podemos escribir **ejemplo3**. Ahora tenemos que poner en el recuadro **Expresión numérica** la función **IDF.CHISQ(p,gl)** que calcula el valor de  $k$  tal que  $P(X \leq k) = p$  siendo  $X$  una variable aleatoria chi-cuadrado con  $gl$  grados de libertad. Para ello, en **Grupo de funciones** hacemos *click* sobre **GL inversos**; en la lista de **Funciones y variables específicas** hacemos doble *click* sobre **Idf.Chisq**. En **Expresión numérica** aparece lo siguiente: **IDF.CHISQ(?,?)**. Tenemos que sustituir las interrogaciones por lo que queremos; es decir, debe aparecer **IDF.CHISQ(0.98,20)**. Después de pulsar el botón **Aceptar** obtendremos en la carpeta de datos, a continuación de nuestra última variable, los resultados de la probabilidad pedida,  $k = 35'0196$ . Podemos volver a grabar estos nuevos datos en **Probabilidad.sav**.

Si nos dan probabilidades de los tipos  $P(X > a)$ ,  $P(X \geq a)$ ,  $\dots$ , tenemos que utilizar lápiz y papel, y aplicar las propiedades de la probabilidad para llegar a expresiones en las que sólo aparezcan probabilidades del tipo  $P(X \leq x)$  (función de distribución).

Por ejemplo, vamos a calcular el valor  $k$  que verifica  $P(X \geq k) = 0'9$ , cuando  $X \equiv \chi_{20}^2$ .

$$\begin{aligned} P(X \geq k) = 0'9 &\Rightarrow P(\overline{X \geq k}) = 1 - 0'9 \Rightarrow P(X < k) = 0'1 \\ &\Rightarrow P(X \leq k) = 0'1 \text{ (por ser continua la variable)} \\ &\Rightarrow k = IDF.CHISQ(0.1, 20) = 12.4426 \end{aligned}$$

Entre las funciones que aparecen dentro de la opción **GL inversos** del recuadro **Grupo de funciones**, las que calculan los resultados de los percentiles o cuantiles de las variables aleatorias estudiadas son:

**IDF.NORMAL(p,media,desv\_típ)** Calcula el valor de  $a$  tal que  $P(X \leq a) = p$ , siendo  $X$  una variable aleatoria normal de parámetros **media** y **desv\_típ**.

**IDF.CHISQ(p,gl)** Calcula el valor de  $a$  tal que  $P(X \leq a) = p$ , siendo  $X$  una variable aleatoria chi-cuadrado con los grados de libertad **gl**.

**IDF.T(p,gl)** Calcula el valor de  $a$  tal que  $P(X \leq a) = p$ , siendo  $X$  una variable aleatoria  $t$  de Student con los grados de libertad **gl**.

**IDF.F(p,gl1,gl2)** Calcula el valor de  $a$  tal que  $P(X \leq a) = p$ , siendo  $X$  una variable aleatoria  $F$  de Snedecor con los grados de libertad **gl1** y **gl2**.

#### Ejercicios Propuestos

11. Hallar el valor de los siguientes percentiles:

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

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

a)  $P(Z \leq a) = 0'725$ .

b)  $P(Z \geq a) = 0'875$ .

13. Si  $X$  es una variable aleatoria con distribución Normal de media  $0'5$  y desviación típica  $0'25$ , determinar el valor de  $a$  tal que:

a)  $P(X < a) = 0'725$ .

b)  $P(X > a) = 0'875$ .

14. Calcular el valor de los siguientes percentiles:

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

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

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

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

16. Calcular el valor de los siguientes percentiles:

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

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

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

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

18. Calcular el valor de los siguientes percentiles:

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

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

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

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

## 5. Contrastes paramétricos

A partir de ahora, volvemos a trabajar con el archivo **Libros.sav**, salvo que se especifique otro archivo para determinados procedimientos.

### 5.1. Contraste para una media. Intervalo de confianza para la media

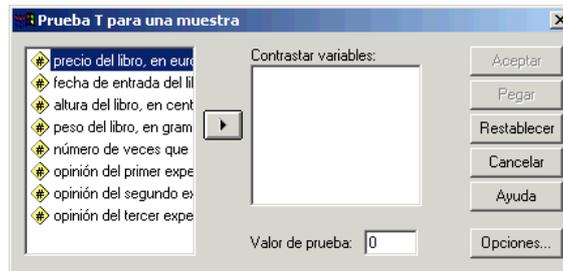


Figura 94: Cuadro de diálogo para hacer un test de hipótesis sobre una media

Para realizar un contraste de hipótesis sobre una media se ha de seleccionar **Analizar** ⇒ **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).

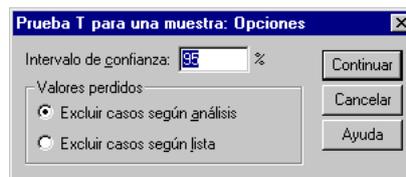


Figura 95: Opciones de un test de hipótesis sobre una media poblacional

En el recuadro **Contrastar Variables** del cuadro de diálogo que aparece (Figura 94) 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 95. 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 96: Estadísticos en el contraste sobre la media de la variable préstamo

	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 97: Resultados del contraste sobre la media de la variable préstamo

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que el número medio anual de préstamos, en la población de todos los libros de la biblioteca, es igual a 14. Lo que queremos es hacer un contraste sobre la media poblacional de la variable **préstamo**.

La hipótesis nula es que la media poblacional de la variable **préstamo** es igual a 14. Como el tamaño muestral de nuestro archivo **Libros.sav** es igual a 30, se puede aplicar esta técnica a cualquier variable (concretamente, se le puede aplicar a la variable **préstamo**). Vamos a contrastar la hipótesis nula  $H_0 : \mu = 14$  frente a la hipótesis alternativa  $H_1 : \mu \neq 14$ . Para ello, seleccionamos **Analizar**  $\Rightarrow$  **Comparar medias**  $\Rightarrow$  **Prueba T para una muestra**. En el recuadro **Contrastar Variables** del cuadro de diálogo que aparece (Figura 94) ponemos la variable **préstamo** (haciendo *click* sobre su nombre o etiqueta y *click* sobre el botón ); en **Valor de prueba** escribimos **14** y pulsamos en **Aceptar**. Los resultados que aparecen en la *ventana del Visor SPSS* son los de las Figuras 96 y 97.

Como el nivel crítico, *Sig. (bilateral)*, es igual a 0'824, 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:

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

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:

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

## 5.2. 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 98.

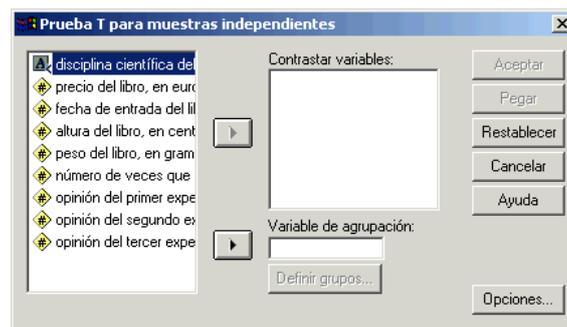


Figura 98: Ventana para hacer un test de hipótesis sobre comparación de dos medias con muestras independientes

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 99: Ventana para definir los grupos en una comparación de dos medias con muestras independientes

En **Contrastar variables** (Figura 98) 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 (que tenga sólo dos resultados distintos) 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 99).

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 99) y se designan los dos grupos (**Grupo 1** y **Grupo 2**) 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 99) si exceptuamos **Punto de corte**. Si la variable es ya dicotómica se designan los dos grupos (**Grupo 1** y **Grupo 2**) 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 98 también tenemos la posibilidad de activar **Opciones**, cuya interpretación es completamente análoga a la del contraste mencionado (ver Figura 95).

**Estadísticos de grupo**

	PRÉSTAMO	N	Media	Desviación típ.	Error típ. de la media
PRECIO	>= 12	15	59,20	25,152	6,494
	< 12	15	48,93	16,020	4,136

**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	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
PRECIO	Se han asumido varianzas iguales	2,078	,161	1,333	28	,193	10,27	7,700	-5,505	26,038
	No se han asumido varianzas iguales			1,333	23,754	,195	10,27	7,700	-5,633	26,166

Figura 100: Resultados del contraste sobre comparación de dos medias con muestras independientes

Con los datos del archivo **Libros.sav**, vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05,

que la media (poblacional) del precio de los libros que se prestan poco es igual a la media (poblacional) del precio de los libros que se prestan mucho. Lo que queremos hacer es un contraste de comparación de dos medias poblacionales (media del precio en el grupo de libros que se prestan poco y media del precio en el grupo de libros que se prestan mucho). Las muestras son independientes, pues la muestra de libros que se prestan poco es independiente de la muestra de libros que se prestan mucho.

Recordemos que los resultados de la variable **préstamo** están comprendidos entre 1 y 42, y que su mediana es 11'5. Podemos dicotomizar la variable **préstamo** considerando que los libros que se prestan poco son los que tienen un resultado menor que 12 y los que se prestan mucho son los que tienen un resultado mayor o igual que 12.

Por tanto, en **Contrastar variables** (de la Figura 98) colocamos **precio** (haciendo *clic* sobre su nombre o etiqueta y *clic* sobre el botón ) y en **Variable de agrupación** seleccionamos **préstamo** (haciendo *clic* sobre su nombre o etiqueta y *clic* sobre el botón ). Pulsamos, a continuación, el botón **Definir grupos**, y en el cuadro que resulta (Figura 99) activamos la opción **Punto de corte** y escribimos en el recuadro el valor 12 (un grupo está formado por los valores menores que 12, y el otro por el resto). Pulsamos en **Continuar** y después en **Aceptar**. En la *ventana del Visor SPSS* aparecen los resultados que se muestran en la Figura 100.

El tamaño muestral del grupo 1 (resultado de **préstamo** mayor o igual que 12) es de 15, y el tamaño muestral del grupo 2 (resultado de **préstamo** menor que 12) es también de 15. Ninguno de estos tamaños son grandes (mayores o iguales que 30), pero suponemos que la variable **precio** es Normal (consultar el apartado 6.5 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 primera tabla de la Figura 100 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 2'078 y el nivel crítico (*Sig.*) es 0'161. Como este nivel crítico es mayor que el nivel de significación,  $\alpha = 0'05$ , 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:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = 1'333,$$

los grados de libertad son:

$$gl = 28,$$

el nivel crítico o p-valor es:

$$\text{Sig. (bilateral)} = 0'193.$$

Como dicho nivel crítico es mayor que el nivel de significación,  $\alpha = 0'05$ , debemos aceptar la hipótesis nula; es decir, la media del precio de los libros que se prestan poco es igual a la media del precio de los libros que se prestan mucho. 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):

$$\text{Diferencia de medias} = 10'27,$$

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

$$7'7,$$

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

$$-5'505 < \mu_1 - \mu_2 < 26'038.$$

### 5.2.1. 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. Vamos a considerar un nivel de significación de  $\alpha = 0'05$ . En primer lugar tenemos que introducir estos datos en el editor y luego grabarlos en un nuevo archivo. 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**  $\Rightarrow$  **Nuevo**  $\Rightarrow$  **Datos**. En segundo lugar definimos convenientemente las dos variables. A continuación introducimos los datos de la manera indicada (ver la Figura 101). Y por último grabamos estos datos en un nuevo archivo llamado **LongitudResumenes.sav**

	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
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
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
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 101: Datos del archivo LongitudResumenes.sav

Ahora ya podemos, con los datos del archivo **LongitudResumenes.sav**, realizar el contraste mencionado. Para ello, en el cuadro de diálogo de la Figura 98 se selecciona la variable **longitud** en el recuadro **Contrastar variables**; y la variable **grupo** en el recuadro **Variable de agrupación**. Se pulsa en **Definir grupos** y en el cuadro de diálogo de la Figura 99 se activa **Usar valores especificados**, y se coloca un **1** junto a **Grupo 1** y un **2** junto a **Grupo 2**. Se pulsa en **Continuar** y después en **Aceptar**. Los resultados de este contraste se muestran en la tabla de la Figura 102.

Observamos que, en el contraste de comparación de varianzas poblacionales, el nivel crítico (*Sig.*) es igual a 0'003, menor que el nivel de significación,  $\alpha = 0'05$ , 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 el nivel de significación,  $\alpha = 0'05$ , por lo que rechazamos la hipótesis de igualdad de medias. Por tanto, aceptamos que la media (poblacional) del número de palabras por resumen escrito en francés es distinta de la media (poblacional) 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; es decir, la media (poblacional) del número de palabras por resumen escrito en francés es menor que la media (poblacional) del número de palabras por resumen escrito en inglés.

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

Volvemos a trabajar con el archivo **Libros.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 102: Resultados del contraste sobre comparación de las medias de las longitudes de los resúmenes

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 103.

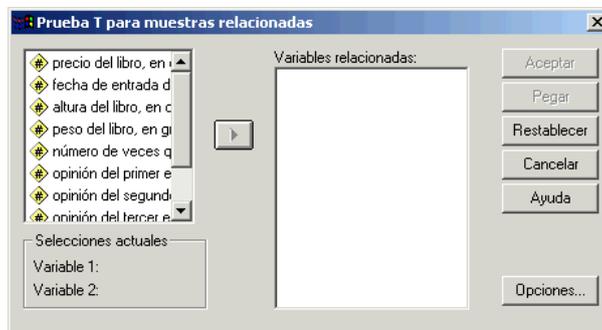


Figura 103: Ventana para hacer un test de hipótesis sobre comparación de dos medias con muestras relacionadas

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 (relacionadas o dependientes). Recordemos que para aplicar esta técnica 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 (hacer clic sobre) 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 103 también tenemos la posibilidad de activar **Opciones**, cuya interpretación es completamente análoga a la del contraste mencionado (ver Figura 95).

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que la media (poblacional) de los resultados de la opinión del primer experto en encuadernación es igual a la media (poblacional) de los resultados de la opinión del segundo experto en encuadernación. Lo que queremos es hacer un contraste de comparación de las medias poblacionales de las variables **opinión1** y **opinión2**. Las muestras son apareadas (o relacionadas) pues realmente las muestras están constituidas por los mismos individuos (libros) a los cuales se les *observa* ambas variables.

Como el tamaño muestral del archivo de datos **Libros.sav** es 30, podemos aplicar este contraste a las variables **opinión1** y **opinión2**. Los resultados obtenidos se muestran en la Figura 104.

En la primera tabla de la Figura 104 aparecen los resultados de los estadísticos para cada una de las variables. Las

		Media	N	Desviación típ.	Error típ. de la media
Par 1	OPINIÓN1	2,93	30	1,413	,258
	OPINIÓN2	3,17	30	1,206	,220

		N	Correlación	Sig.
Par 1	OPINIÓN1 y OPINIÓN2	30	,007	,972

		Diferencias relacionadas					t	gl	Sig. (bilateral)
		Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	OPINIÓN1 - OPINIÓN2	-.23	1,851	,338	-.92	,46	-.690	29	,495

Figura 104: Resultados del contraste sobre comparación de dos medias con muestras relacionadas

medias muestrales son  $\bar{x}_1 = 2'93$  y  $\bar{x}_2 = 3'17$ ; las cuasidesviaciones típicas muestrales son  $S_1 = 1'413$  y  $S_2 = 1'206$ ; los errores típicos de las medias muestrales son  $S_1/\sqrt{n} = 0'258$  y  $S_2/\sqrt{n} = 0'220$ , respectivamente.

En la segunda tabla de la Figura 104 aparecen los resultados de la correlación entre las dos variables. El valor del coeficiente de correlación lineal de Pearson en la muestra es  $r = 0'007$ . 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  $0'972$ , mayor que el nivel de significación,  $\alpha = 0'05$ , por lo que podemos aceptar la hipótesis nula. Por tanto, aceptamos que las dos variables, **opinión1** y **opinión2**, son independientes en la población de todos los libros.

En la tercera tabla de la Figura 104 aparecen los siguientes resultados: la media de la variable diferencia  $D = X_1 - X_2$  es igual a:

$$\text{Media} = \bar{d} = -0'23,$$

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

$$\text{Desviación típ.} = S_d = 1'851,$$

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

$$\text{Error típ. de la media} = \frac{S_d}{\sqrt{n}} = 0'338,$$

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

$$-0'92 < \mu_1 - \mu_2 < 0'46,$$

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

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = -0'690,$$

los grados de libertad son:

$$gl = 29,$$

y el nivel crítico es:

$$\text{Sig. (bilateral)} = 0'495.$$

Como dicho nivel crítico es mayor que el nivel de significación,  $\alpha = 0'05$ , 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 (poblacionales) de las opiniones de los dos primeros expertos en encuadernación.

## 6. Contrastes no paramétricos

### 6.1. Contraste sobre una proporción

Para hacer un contraste de hipótesis sobre una proporción poblacional se selecciona **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **Binomial**. Si la variable estudiada no es dicotómica (no tiene sólo dos resultados distintos), debe dicotomizarse como veremos a continuación. Se especifican, como de costumbre, las variables a tratar en **Contrastar variables** teniendo presente

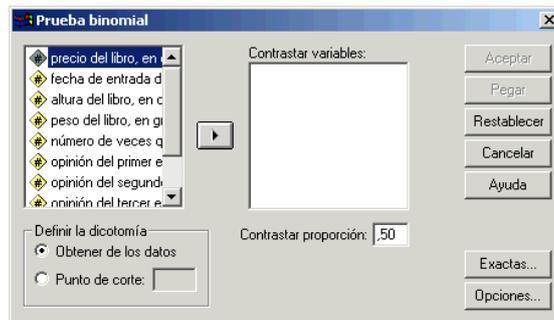


Figura 105: Cuadro de diálogo para el contraste sobre una proporción

que no tienen por qué ser originariamente dicotómicas (Figura 105). 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.

Con los datos de nuestro archivo **Libros.sav**, 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. Vamos a considerar un nivel de significación de  $\alpha = 0'05$ .

Para hacer este test de hipótesis, en **Contrastar variables** (Figura 105) 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 106. 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, menor que el nivel de significación,  $\alpha = 0'05$ , 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 106: Contraste sobre una proporción para la variable préstamo

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

Para realizar este contraste se selecciona **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **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 107) 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.

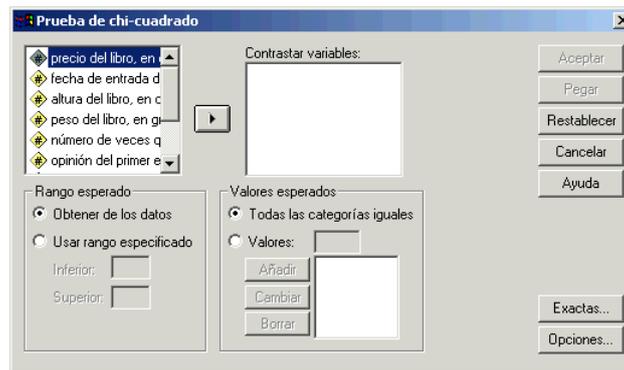


Figura 107: Cuadro de diálogo para el contraste chi-cuadrado sobre proporciones

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.

**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 los porcentajes o las frecuencias relativas (proporciones) esperadas. 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 absolutas 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.

Con los datos de nuestro archivo **Libros.sav**, vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que las proporciones en que salen los resultados 1, 2, 3, 4 y 5 de la variable **opinión1** en toda la población de libros de la biblioteca son, respectivamente, 0'2, 0'25, 0'15, 0'25 y 0'15; es decir, si  $X$  es la variable aleatoria **opinión1** se quiere comprobar si se cumple:

$$P(X = 1) = 0'2, \quad P(X = 2) = 0'25, \quad P(X = 3) = 0'15, \quad P(X = 4) = 0'25, \quad P(X = 5) = 0'15$$

Para ello, en el recuadro **Contrastar variables** del cuadro de diálogo inicial (Figura 107) se selecciona la variable **opinión1**. En el grupo **Rango esperado** activamos **Obtener de los datos**. En el grupo **Valores esperados** seleccionamos **Valores**. A continuación debemos especificar las frecuencias (o los porcentajes o las proporciones) esperadas. Si queremos especificar las proporciones (probabilidades) esperadas, hacemos lo siguiente:

- En **Valores** escribimos 0,2 y pulsamos **Añadir**. Esto significa que la proporción (o probabilidad) esperada para el primer resultado de la variable (que es 1) es igual a 0,2 (por tanto, el porcentaje esperado es el 20 %).
- En **Valores** escribimos 0,25 y pulsamos **Añadir**. Esto significa que la proporción (o probabilidad) esperada para el segundo resultado de la variable (que es 2) es igual a 0,25 (por tanto, el porcentaje esperado es el 25 %).
- En **Valores** escribimos 0,15 y pulsamos **Añadir**. Esto significa que la proporción (o probabilidad) esperada para el tercer resultado de la variable (que es 3) es igual a 0,15 (por tanto, el porcentaje esperado es el 15 %).
- En **Valores** escribimos 0,25 y pulsamos **Añadir**. Esto significa que la proporción (o probabilidad) esperada para el cuarto resultado de la variable (que es 4) es igual a 0,25 (por tanto, el porcentaje esperado es el 25 %).
- En **Valores** escribimos 0,15 y pulsamos **Añadir**. Esto significa que la proporción (o probabilidad) esperada para el quinto resultado de la variable (que es 5) es igual a 0,15 (por tanto, el porcentaje esperado es el 15 %).

Si se quiere especificar los porcentajes esperados, hacemos lo mismo pero en vez de escribir 0, 2; 0, 25; 0, 15; 0'25 y 0, 15 se escribe 20, 25, 15, 25 y 15. Por último, si se quiere especificar las frecuencias absolutas esperadas, tecleamos los siguientes valores: 6; 7, 5; 4, 5; 7, 5 y 4, 5 (pues 6 es el 20 % de 30, 7'5 es el 25 % de 30 y 4'5 es el 15 % de 30).

Los resultados se muestran en la Figura 108. Como el nivel crítico (*Sig. asintót.*) es igual a 0'996, mayor que el nivel de significación,  $\alpha = 0'05$ , 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 40 % de frecuencias esperadas menores que 5.

**OPINIÓN1**

	N observado	N esperado	Residual
1	6	6,0	,0
2	7	7,5	-,5
3	5	4,5	,5
4	7	7,5	-,5
5	5	4,5	,5
Total	30		

**Estadísticos de contraste**

	OPINIÓN1
Chi-cuadrado <sup>a</sup>	,178
gl	4
Sig. asintót.	,996

a. 2 casillas (40,0%) tienen frecuencias esperadas menores que 5. La frecuencia de casilla esperada mínima es 4,5.

Figura 108: Resultados del contraste chi-cuadrado sobre proporciones para la variable opinión1

### 6.3. 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 109: Cuadro de diálogo de la opción Tablas de contingencia

Como se ve en la Figura 109, 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 *clic* 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 la *ventana del Visor SPSS* las tablas de las Figuras 110 y 111.

En el cuadro de diálogo de la Figura 109 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

	Casos					
	Válidos		Perdidos		Total	
	N	%	N	%	N	%
OPINIÓ1 * OPINIÓ2	30	100%	0	,0%	30	100%

Figura 110: Tabla-resumen de casos de las variables opinión1 y opinión2

		OPINIÓ2					Total
		1	2	3	4	5	
OPINIÓ1	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 111: Tabla de contingencia de las variables opinión1 y opinión2

especificadas en **Filas** y **Columnas**. Es posible especificar más de una variable de control. En el caso de que hubiese, por 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** (Figura 109) seleccionamos la variable **opinión1**, en **Columnas** la variable **opinión2** y en el recuadro de las variables de control seleccionamos la variable **materia**, obtenemos, en la *ventana del Visor SPSS*, la tabla de la Figura 112.

MATERIA			OPINIÓ2					Total
			1	2	3	4	5	
E	OPINIÓ1	1			1			1
		2	1			1	2	4
		3		1				1
		4			1		1	2
	Total		1	1	2	1	3	8
G	OPINIÓ1	1		1				1
		2			1			1
		3			1			1
		5		1	1	2		4
	Total			2	3	2		7
H	OPINIÓ1	1			1			1
		3		1	1			2
		4	1					1
	Total		1	1	2			4
I	OPINIÓ1	1		1	1		1	3
		2			1		1	2
		5					1	1
	Total			1	2		3	6
L	OPINIÓ1	3			1			1
		4		1	1	1		3
	Total			1	2	1		4

Figura 112: Tabla de contingencia de opinión1 y opinión2 con la variable control materia

En el cuadro de diálogo de la Figura 109 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 109 nos quedan tres botones por explicar: **Estadísticos**, **Casillas** y **Formato**.

Si pulsamos el botón **Estadísticos** nos aparece el cuadro de diálogo de la Figura 113. En este cuadro se pueden elegir los coeficientes de correlación que deseamos 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

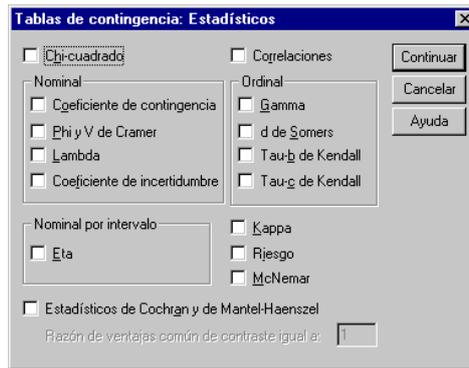


Figura 113: Cuadro de diálogo de la opción Estadísticos

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, con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar la hipótesis de independencia de las variables **opinión1** y **opinión2** en la población de todos los libros de la biblioteca, con un nivel de significación de 0'05.

Para ello, en **Filas** (Figura 109) seleccionamos la variable **opinión1**, en **Columnas** la variable **opinión2** y al pulsar el botón **Estadísticos** seleccionamos **Chi-cuadrado**. Obtenemos entonces las tablas de las Figuras 110, 111 y 114.

	Valor	gl	Sig. asintótica (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 114: 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 el nivel de significación,  $\alpha = 0'05$ , por lo que deberíamos aceptar la hipótesis nula de independencia de las variables **opinión1** y **opinión2** en la población de todos los libros de la biblioteca. 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 109 nos aparece el nuevo cuadro de diálogo de la Figura 115.

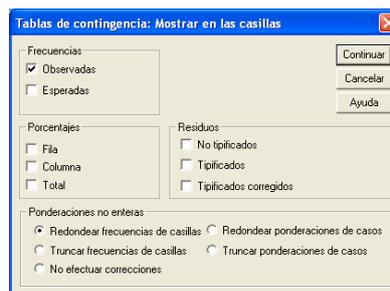


Figura 115: Cuadro de diálogo para especificar más cosas sobre las casillas de las tablas de contingencia

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.

**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.

En el bloque **Ponderaciones no enteras** se nos presentan varias opciones que se refieren a lo siguiente: Los recuentos de las casillas suelen ser valores enteros, ya que representan el número de casos de cada casilla; sin embargo, si el archivo de datos está ponderado en un momento determinado por una variable de ponderación con valores fraccionarios (por ejemplo, 1/25), los recuentos de las casillas pueden que también sean valores fraccionarios, caso en que se pueden redondear estos valores antes o después de calcular los recuentos de las casillas o bien utilizar recuentos de casillas fraccionarios en la presentación de las tablas y los cálculos de los estadísticos.

Por ejemplo, en el cuadro de diálogo de la Figura 109 en **Filas** seleccionamos la variable **opinión1**, en **Columnas** la variable **opinión2** y pulsamos en **Casillas**. En el cuadro de diálogo de la Figura 115 en **Frecuencias** seleccionamos **Observadas** y **Esperadas**; en **Porcentajes** seleccionamos **Total**; y en **Residuos** seleccionamos **No tipificados**. Entonces obtenemos los resultados de la Figura 116.

**Tabla de contingencia OPINIÓN1 \* OPINIÓN2**

			OPINIÓN2					Total
			1	2	3	4	5	
OPINIÓN1	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%	,0%	3,3%	20,0%
	2	Residuo	-,4	,6	,8	-,8	-,2	
		Recuento	1	0	2	1	3	7
		Frecuencia esperada	,5	1,6	2,6	,9	1,4	7,0
	3	% del total	3,3%	,0%	6,7%	3,3%	10,0%	23,3%
		Residuo	,5	-1,6	-,6	,1	1,6	
		Recuento	0	2	3	0	0	5
	4	Frecuencia esperada	,3	1,2	1,8	,7	1,0	5,0
		% del total	,0%	6,7%	10,0%	,0%	,0%	16,7%
		Residuo	-,3	,8	1,2	-,7	-1,0	
	5	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%
Total	Residuo	,5	,4	-,6	,1	-,4		
	Recuento	0	1	1	2	1	5	
	Frecuencia esperada	,3	1,2	1,8	,7	1,0	5,0	
Total	% del total	,0%	3,3%	3,3%	6,7%	3,3%	16,7%	
	Residuo	-,3	-,2	-,8	1,3	,0		
	Recuento	2	7	11	4	6	30	
Total	Frecuencia esperada	2,0	7,0	11,0	4,0	6,0	30,0	
	% del total	6,7%	23,3%	36,7%	13,3%	20,0%	100,0%	

Figura 116: Nueva tabla de contingencia de las variables opinión1 y opinión2

Si pulsamos el botón **Formato** del cuadro de diálogo de la Figura 109 nos aparece el nuevo cuadro de diálogo de la Figura 117. En este cuadro sólo es posible reordenar las filas:

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

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

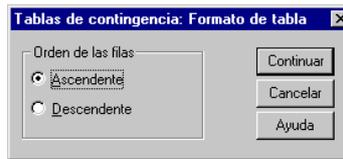


Figura 117: Cuadro de diálogo para especificar el formato de las tablas de contingencia

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

Supongamos que queremos comprobar si se puede aceptar, con un nivel de significación de 0'05, que 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

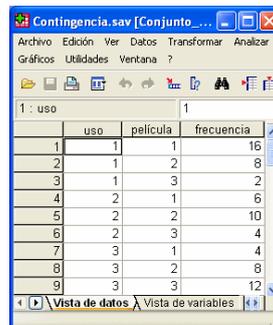


Figura 118: Datos del archivo Contingencia.sav

En primer lugar tenemos que introducir estos datos. Para ello, se selecciona **Archivo** ⇒ **Nuevo** ⇒ **Datos**. 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 **frecuencia** 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 118. No olvidemos grabar ahora este archivo 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 119: 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 hay que activar la opción **Ponderar casos mediante** y hay que seleccionar la variable **frecuencia** en el recuadro **Variable de ponderación**.

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 109 seleccionamos la variable **uso** en **Filas** y la variable **película** en **Columnas**. Pulsamos el botón **Estadísticos** y en el cuadro que aparece (Figura 113) activamos **Chi-cuadrado**. En el cuadro de diálogo de la Figura 109 podemos también pulsar el botón **Casillas** y activar (Figura 115) las opciones **Esperadas** (Frecuencias) y **No tipificados** (Residuos). Los resultados se muestran en las Figuras 119 y 120.

**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 120: Contraste chi-cuadrado sobre independencia entre las variables uso y película

Según los resultados (Figura 120), el nivel crítico (*Sig. asint. (bilateral)*) es igual a 0'002, menor que el nivel de significación,  $\alpha = 0'05$ , por lo que rechazamos la hipótesis nula de que las variables **uso** y **película** son independientes. Por tanto, aceptamos que dichas variables son dependientes.

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

Volvemos a trabajar con el archivo **Libros.sav**.

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



Figura 121: Cuadro de diálogo para el contraste de las rachas sobre aleatoriedad de la muestra

**Contrastar variables** (Figura 121) se declaran las variables a analizar. Si hacemos *click* 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.

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que la muestra de resultados de la variable **precio** es aleatoria. Para ello, seleccionamos **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **Rachas**. En **Contrastar variables** ponemos la variable **precio** y pulsamos en **Aceptar**.

La conclusión (Figura 122) 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 0'353, mayor que el nivel de significación,  $\alpha = 0'05$ .

#### 6.5. 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 123). Como siempre, en **Contrastar variables** son declaradas las variables que individualmente serán sometidas a análisis. Por defecto, haciendo *click* en **Aceptar**, la distribución con la que se compara la observada es la Normal definida por la misma media y desviación

	PRECIO
Valor de prueba <sup>a</sup>	47,00
Casos < Valor de prueba	15
Casos >= Valor de prueba	15
Casos en total	30
Número de rachas	13
Z	-,929
Sig. asintót. (bilateral)	,353

a. Mediana

Figura 122: Contraste de las rachas sobre aleatoriedad de los resultados muestrales de la variable precio

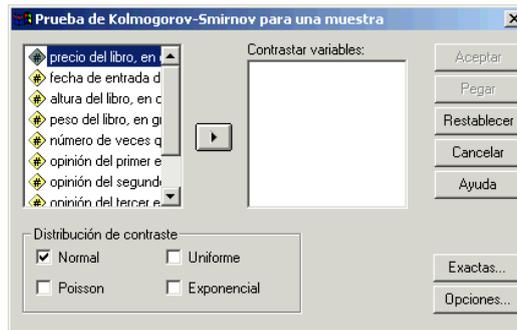


Figura 123: Cuadro de diálogo para el contraste de Kolmogorov–Smirnov sobre bondad de ajuste

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**.

		PRECIO	ALTURA	PESO	PRÉSTAMO
N		30	30	30	30
Parámetros normales <sup>a,b</sup>	Media	54,07	23,820	580,17	13,60
	Desviación típica	21,367	5,0531	184,935	9,737
Diferencias más extremas	Absoluta	,160	,212	,128	,169
	Positiva	,160	,212	,128	,169
	Negativa	-,130	-,180	-,127	-,098
Z de Kolmogorov-Smirnov		,878	1,159	,699	,927
Sig. asintót. (bilateral)		,424	,136	,714	,356

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Figura 124: Ejemplo del contraste de Kolmogorov–Smirnov sobre bondad de ajuste

Con los datos del archivo **Libros.sav**, veamos si se puede aceptar, con un nivel de significación de 0'05, que las variables **precio**, **altura**, **peso** y **préstamo** siguen una distribución Normal (en la población de todos los libros de la biblioteca). A la vista de los resultados (Figura 124) se puede aceptar que las cuatro variables son normales, ya que el nivel crítico (*Sig. asintót. bilateral*) es, en los cuatro casos, mayor que el nivel de significación,  $\alpha = 0'05$ .

## 6.6. Contraste de homogeneidad con 2 muestras independientes

En general, el procedimiento que vamos a explicar se aplica cuando la variable es cuantitativa o cualitativa ordinal pero no es Normal y, además, los tamaños muestrales no son grandes; caso en el que no se puede usar la técnica del apartado 5.2.

El cuadro de diálogo inicial se logra con **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **2 muestras independientes** (Figura 125). En **Contrastar variables** se especifica la variable (o variables) objeto de estudio. Si se especifica más de una variable, éstas serán analizadas individualmente. En el recuadro **Variable de agrupación** se ha de especificar una variable (con resultados numéricos) 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 126). En el grupo **Tipo de prueba** se presentan, además del contraste **U de Mann–Whitney** ya conocido, otros como: **Reacciones extremas de Moses**, **Z de Kolmogorov–Smirnov** y **Rachas de Wald–Wolfowitz**.

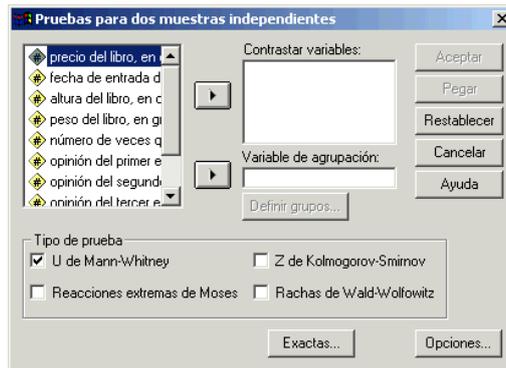


Figura 125: Cuadro de diálogo para el contraste de homogeneidad con dos muestras independientes



Figura 126: Ventana para definir los grupos en un contraste de homogeneidad con dos muestras independientes

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que la distribución de la variable **precio** es la misma en el grupo en que la variable **opinión1** vale 2 (“mala encuadernación” según el criterio del primer experto) y en el grupo en que la variable **opinión1** vale 4 (“buena encuadernación” según el criterio del primer experto).

Para ello, en **Contrastar variables** (Figura 125) colocamos la variable **precio** y en el recuadro **Variable de agrupación** elegimos la variable **opinión1**. Pulsamos **Definir grupos** y, en el cuadro que aparece (Figura 126), escribimos **2** en **Grupo 1** y **4** 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 127. Como el nivel crítico (*Sig. asintót. bilateral*) es igual a 0'063, mayor que el nivel de significación,  $\alpha = 0'05$ , entonces podemos aceptar la hipótesis nula de homogeneidad; es decir, el precio es el mismo en el grupo de libros cuya encuadernación es mala (según el criterio del primer experto) que en el grupo de libros cuya encuadernación es buena (según el criterio del primer experto).

**Rangos**

	OPINIÓN1	N	Rango promedio	Suma de rangos
PRECIO	2	7	9,57	67,00
	4	7	5,43	38,00
	Total	14		

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

	PRECIO
U de Mann-Whitney	10,000
W de Wilcoxon	38,000
Z	-1,857
Sig. asintót. (bilateral)	,063
Sig. exacta [2*(Sig. unilateral)]	,073 <sup>a</sup>

a. No corregidos para los empates.

b. Variable de agrupación: OPINIÓN1

Figura 127: 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 5.2.1: 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.

## 6.7. 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 128). En

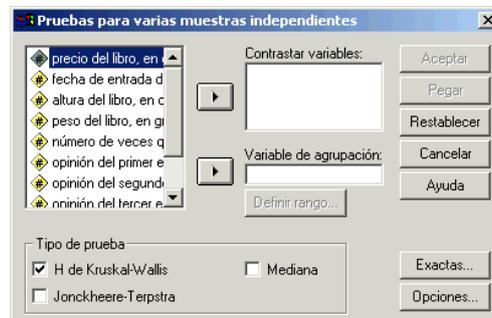


Figura 128: Cuadro de diálogo para el contraste de homogeneidad con más de dos muestras independientes

**Contrastar variables** se especifica la variable (o variables) objeto de estudio. Si se especifica más de una variable, éstas serán analizadas individualmente. Se comparan 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 rango** para especificar los grupos que queremos comparar (Figura 129).

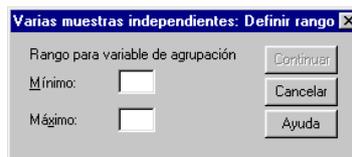


Figura 129: Ventana para definir los grupos en un contraste de homogeneidad con más de dos muestras independientes

En el grupo **Tipo de prueba** (Figura 128) se presenta, además del conocido contraste **H de Kruskal-Wallis**, el test de la **Mediana** y el test de **Jonckheere-Terpstra**.

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que la distribución de la variable **precio** es la misma en los cinco grupos en que tenemos dividida la variable **opinión1** (1=muy mala, 2=mala, 3=regular, 4=buena, 5=muy buena).

Para ello, en **Contrastar variables** colocamos la variable **precio** y en el recuadro **Variable de agrupación** elegimos la variable **opinión1**. Pulsamos **Definir rango** y, en el cuadro que aparece (Figura 129), escribimos **1** en **Mínimo** y **5** 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 130. Como el nivel crítico (*Sig. asintót.*) es igual a 0'745, mayor que el nivel de significación,  $\alpha = 0'05$ , aceptamos la hipótesis nula de homogeneidad; es decir, la distribución de la variable **precio** es la misma en las cinco categorías; es decir, el precio no depende de la calidad de la encuadernación de los libros.

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 5.2.1: 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.

## 6.8. Contraste de homogeneidad con 2 muestras relacionadas

En general, el procedimiento que vamos a explicar se aplica cuando las variables son cuantitativas o cualitativas ordinales pero no son Normales y, además, los tamaños muestrales no son grandes; caso en el que no se puede usar la técnica del apartado 5.3.

El cuadro de diálogo inicial de este procedimiento se logra con **Analizar** ⇒ **Pruebas no paramétricas** ⇒ **2 muestras relacionadas** (Figura 131). En primer lugar debemos seleccionar (hacer *click* sobre) 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 conocido contraste de homogeneidad de **Wilcoxon**, el contraste de los **Signos**, el test de **McNemar** y el contraste de **Homogeneidad marginal**.

**Rangos**

	OPINIÓ1	N	Rango promedio
PRECIO	1	6	16,42
	2	7	19,00
	3	5	13,60
	4	7	13,14
	5	5	14,70
	Total	30	

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

	PRECIO
Chi-cuadrado	1,951
gl	4
Sig. asintót.	,745

- a. Prueba de Kruskal-Wallis
- b. Variable de agrupación: OPINIÓ1

Figura 130: Ejemplo de un contraste de homogeneidad con más de dos muestras independientes

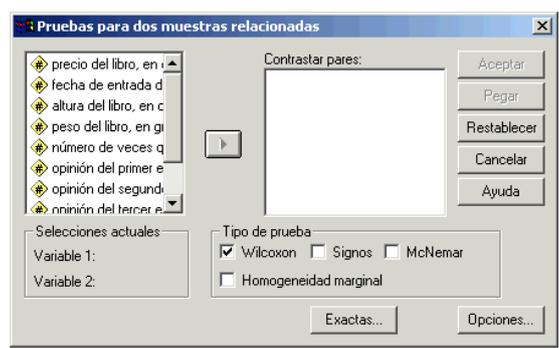


Figura 131: Cuadro de diálogo para el contraste de homogeneidad con dos muestras relacionadas

**Rangos**

		N	Rango promedio	Suma de rangos
OPINIÓ3 - OPINIÓ2	Rangos negativos	10 <sup>a</sup>	9,15	91,50
	Rangos positivos	13 <sup>b</sup>	14,19	184,50
	Empates	7 <sup>c</sup>		
	Total	30		

- a. OPINIÓ3 < OPINIÓ2
- b. OPINIÓ3 > OPINIÓ2
- c. OPINIÓ2 = OPINIÓ3

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

	OPINIÓ3 - OPINIÓ2
Z	-1,447 <sup>a</sup>
Sig. asintót. (bilateral)	,148

- a. Basado en los rangos negativos.
- b. Prueba de los rangos con signo de Wilcoxon

Figura 132: Ejemplo del contraste de Wilcoxon sobre homogeneidad con dos muestras dependientes

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que 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).

En el cuadro de diálogo de la Figura 131 hacemos *clic* sobre el nombre (o las etiquetas) de las variables **opinión2** y **opinión3**, pulsamos el botón  y dejamos seleccionada la opción **Wilcoxon**. Tras pulsar **Aceptar** obtenemos los resultados de la Figura 132. Como el nivel crítico (*Sig. asint. bilateral*) es 0'148, mayor que el nivel de significación,  $\alpha = 0'05$ , 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.

### 6.9. 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 133).

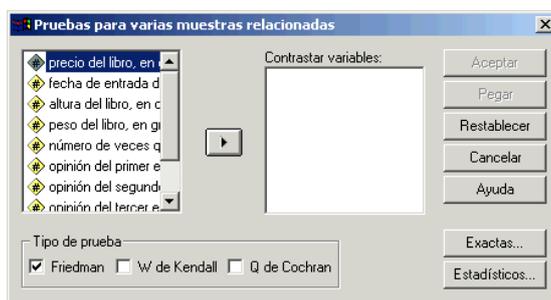


Figura 133: Cuadro de diálogo para el contraste de homogeneidad con más de dos muestras relacionadas

Rangos	
	Rango promedio
OPINIÓN1	1,83
OPINIÓN2	1,97
OPINIÓN3	2,20

Estadísticos de contraste <sup>a</sup>	
N	30
Chi-cuadrado	2,480
gl	2
Sig. asintót.	,289

a. Prueba de Friedman

Figura 134: Ejemplo del contraste de Friedman sobre homogeneidad con más de dos muestras dependientes

Para declarar las variables que serán objeto de análisis debemos seleccionarlas (hacer *clic* sobre su nombre o etiqueta) y pulsar el botón  (lo hacemos una a una). En **Tipo de prueba** se presentan, además del conocido contraste de homogeneidad de **Friedman**, el contraste **W de Kendall** y el **Q de Cochran**.

Con los datos del archivo **Libros.sav** vamos a comprobar si se puede aceptar, con un nivel de significación de 0'05, que la opinión de los tres expertos en la encuadernación de los libros se puede considerar la misma en la población. Notemos que, efectivamente, las tres muestras están relacionadas, pues realmente son la misma muestra (a la cual se le ha observado tres variables distintas).

En el cuadro de diálogo de la Figura 133 marcamos la variable **opinión1** y pulsamos el botón ; marcamos la variable **opinión2** y pulsamos el botón ; y marcamos la variable **opinión3** y pulsamos el botón . Dejamos seleccionada la opción **Friedman** y pulsamos en **Aceptar**. Los resultados aparecen en la Figura 134. Debido a que el nivel crítico (*Sig. asintót.*) es 0'289, mayor que el nivel de significación,  $\alpha = 0'05$ , 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.