

## Presentación general del programa SPSS

### Introducción

El presente curso tiene como objetivo acercar al usuario al manejo del *software* de análisis estadístico SPSS, acrónimo de *Statistical Package for Sciences Socials* (Paquete Estadístico para las Ciencias Sociales), en sus aspectos más básicos, los que se refieren al tratamiento general de datos y los relativos a ciertos análisis estadísticos considerados simples, es decir, descripción general de cualquier tipo de variable estadística y evaluación de relaciones entre dos variables, dejando para un futuro análisis más complejos, de carácter multivariante, que también pueden realizarse con este programa.

En primer lugar, y antes de comenzar a desarrollar los contenidos específicos de este curso, daremos un breve paseo por las versiones anteriores de SPSS para ver la evolución que ha experimentado hasta llegar a la actual versión, la 10.0.

Para el desarrollo del curso se emplean los mismos archivos que SPSS incluye en el CD-ROM en el que se distribuye el programa. En cada momento haremos mención al archivo con el que vamos a trabajar. Todos los archivos, una vez instalado SPSS en el ordenador, se encuentran en la misma ruta C:\Archivos de Programa\SPSS\

Antes de comenzar, expreso el deseo de que este manual os sirva de guía para moveros con sencillez por las pantallas del programa y realizar los procedimientos de análisis más básicos. Por supuesto, aceptaré todos los comentarios que tengáis a bien hacerme para mejorar este manual en la medida de lo posible.

### Breve historia del SPSS

A finales de la década de los 80 SPSS desarrolló un programa de análisis estadístico para su ejecución en los ordenadores personales, bajo el entorno operativo MS-DOS. Hasta entonces había versiones del mismo para grandes plataformas (*mainframe*), que habitualmente conformaban los equipos de los centros de cálculo de las universidades y laboratorios de investigación. Para llevar a cabo los análisis era preciso escribir las instrucciones en un lenguaje específico de SPSS, con una sintaxis particular. Este lenguaje que soportaba SPSS para grandes equipos se ha transmitido, con ligeras variaciones, a las sucesivas versiones para ordenadores personales, tanto en el entorno MS-DOS como en el de WINDOWS, aunque en este último pueda llegar a pasar desapercibido para el principiante.

Como muestra veamos cómo se podría obtener una distribución de frecuencias de una variable V1 contenida en un archivo de datos con tres variables (V1, V2 y V3). Las instrucciones serían las siguientes:

```
DATA LIST FILE ='C:\CURSPSS\ARCHIV1.DAT'/ V1 1-3 V2 5-6 V3 8-20(a).  
FRECUENCIES V1/ STATISTICS = NONE.
```

En términos llanos, estas dos sentencias podrían traducirse así:

*"... leer el archivo de datos en formato ASCII, ARCHIV1.DAT (DATA LIST FILE) ubicado en el directorio CURSPSS de la unidad C, el cual contiene tres variable: V1 con tres dígitos que ocupa las columnas 1 a 3; V2 que ocupan las columnas 5 y 6; y V3 que ocupan las columnas 8 a 20 es una variable de cadena, tal como se especifica por la letra a dentro del paréntesis".*

Posteriormente, confeccionar una distribución de frecuencias de la variable V1, y no calcular estadísticos (STATISTICS = NONE)..."

De esta forma, escribiendo los procedimientos adecuados, se obtenían todos los análisis que incorporaba el SPSS.

Como se ha dicho, esta sintaxis se mantiene, ampliada, en todas las sucesivas versiones que han salido al mercado, para ser implementadas en los ordenadores personales. No obstante, ya en la versión 4 para DOS, aparecieron los primeros menús de ayuda en línea mediante los cuales se podían obtener los mismos resultados sin tener que escribir los procedimientos. De esta forma se elegían en dichos menús los procedimientos que se iban a utilizar y el programa escribía en un editor de texto (REVIEW) la sentencia adecuada en función del procedimiento elegido; SPSS empezaba a dulcificar el *interface* de usuario.

En estas versiones de SPSS para DOS había un déficit importante, que era el asunto de los gráficos. Para obtenerlos era preciso tener grabado en el ordenador algún software de gráficos, y configurar SPSS para que pudiera trabajar con ese software en cuestión (por defecto solía trabajar con HARVARD-GRAPHICS), lo cual, para un usuario poco avezado, podía suponer un problema añadido.

Este inconveniente ha sido subsanado en las versiones para Windows, y SPSS ya dispone de un software de generación de gráficos integrado en la aplicación y con las opciones propias de los editores de gráficos.

Después de este breve repaso por la historia del SPSS vamos a comenzar el curso de la manera más directa posible: realizando una sesión completa de trabajo, que nos permitirá obtener una visión global de las características más notables de la aplicación. Posteriormente, en los siguientes capítulos, profundizaremos en cada una de las operaciones básicas y procedimientos que se pueden realizar, desde la edición de datos a la elaboración de análisis estadísticos, pasando por el tratamiento de esos datos (creación de nuevas variables, transformación de variables, ordenación, ponderación, selección, etc.).

Comencemos pues.

# 1. Primera sesión con SPSS

## 1.1 Aspectos básicos

Cualquier sesión tipo se puede resumir en cuatro grandes apartados:

- Lectura de un conjunto de datos
- Selección del Procedimiento
- Selección de Variables
- Examen de Resultado

Pero antes... antes hemos de entrar en SPSS para poder llevar a cabo esta primera sesión. Para ello hay dos maneras de proceder: 1) Desde el menú Programas que se despliega a pulsar el botón **Inicio** se accede al programa SPSS, de la misma manera que se accede a cualquier programa que opere bajo el sistema operativo de Windows, bien en la versión 95 en la 98 o en la 2000; 2) A través de un Icono de Acceso Directo que hayamos creado previamente en el Escritorio o en la barra de accesos rápidos situada en la parte inferior de la pantalla, por el procedimiento habitual de creación de estos tipos de accesos directos<sup>1</sup>. En ambos casos el resultado es el mismo: se accede al programa, directamente al **Editor de Datos**, cuya apariencia es la que se muestra en la Figura 1.1.

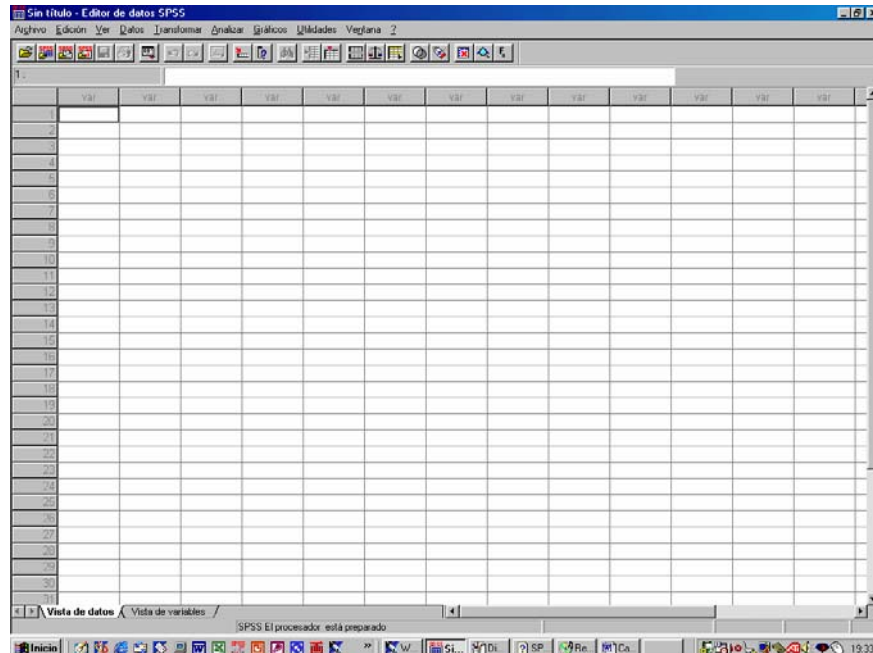



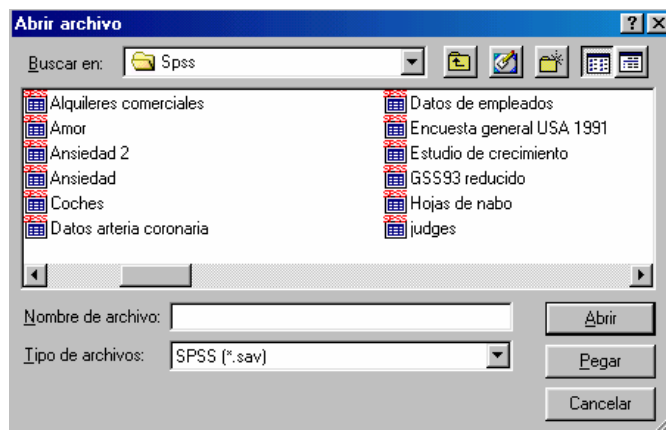
Fig. 1.1 Editor de datos de SPSS, sin datos

<sup>1</sup> La diferencia entre un icono de acceso directo en el escritorio y otro en la barra de acceso rápido está en que en el escritorio, si no se ha modificado las opciones de carpeta del panel de control, hay que hacer doble clic para acceder al programa y en la barra sólo un clic.

## Primera sesión con SPSS

Es en esta pantalla en la que se va a desarrollar buena parte de las sesiones con SPSS. Aquí es donde grabaremos los datos registrados en el desarrollo de nuestros trabajos, o donde se mostrarán los datos ya grabados en archivos cuando queramos someterlos a los procedimientos de análisis de SPSS.

El aspecto del editor de datos es el propio de una rejilla de filas y columnas cuya intersección conforman las celdillas de la misma -cada celdilla un dato-, similar a la que dispone cualquier hoja de cálculo. En esta primera sesión vamos a utilizar los datos previamente almacenados en un archivo, por lo que el primer paso es leer esos datos. Para ello se puede emplear dos maneras alternativas: la primera es a través de la opción **Archivo** del menú principal, sub-opción **Abrir**. La otra alternativa, más inmediata, es pulsar, en los iconos que aparecen debajo del menú general, el correspondiente a **Abrir archivo** . En ambos casos, se accede a una ventana como la de la Figura 1.2.



**Figura 1.2 Cuadro de diálogo de *Abrir archivo***

Por defecto, sólo se lista los archivos de datos generados y guardados previamente por SPSS, que en las versiones para Windows tienen la extensión SAV, aunque SPSS puede leer datos grabados en diferentes formatos (ASCII, dBASE, Excel, etc.), y por supuesto los archivos generados por las anteriores versiones del programa, que se identifican por la extensión SYS.

Para abrir un archivo de datos, basta hacer doble clic con el botón izquierdo del ratón en el mismo y se incorpora al Editor de datos. El aspecto del editor una vez leído el archivo (en este caso el archivo *Datos de empleados*) es el que se ve en la Figura 1.3.



## Primera sesión con SPSS

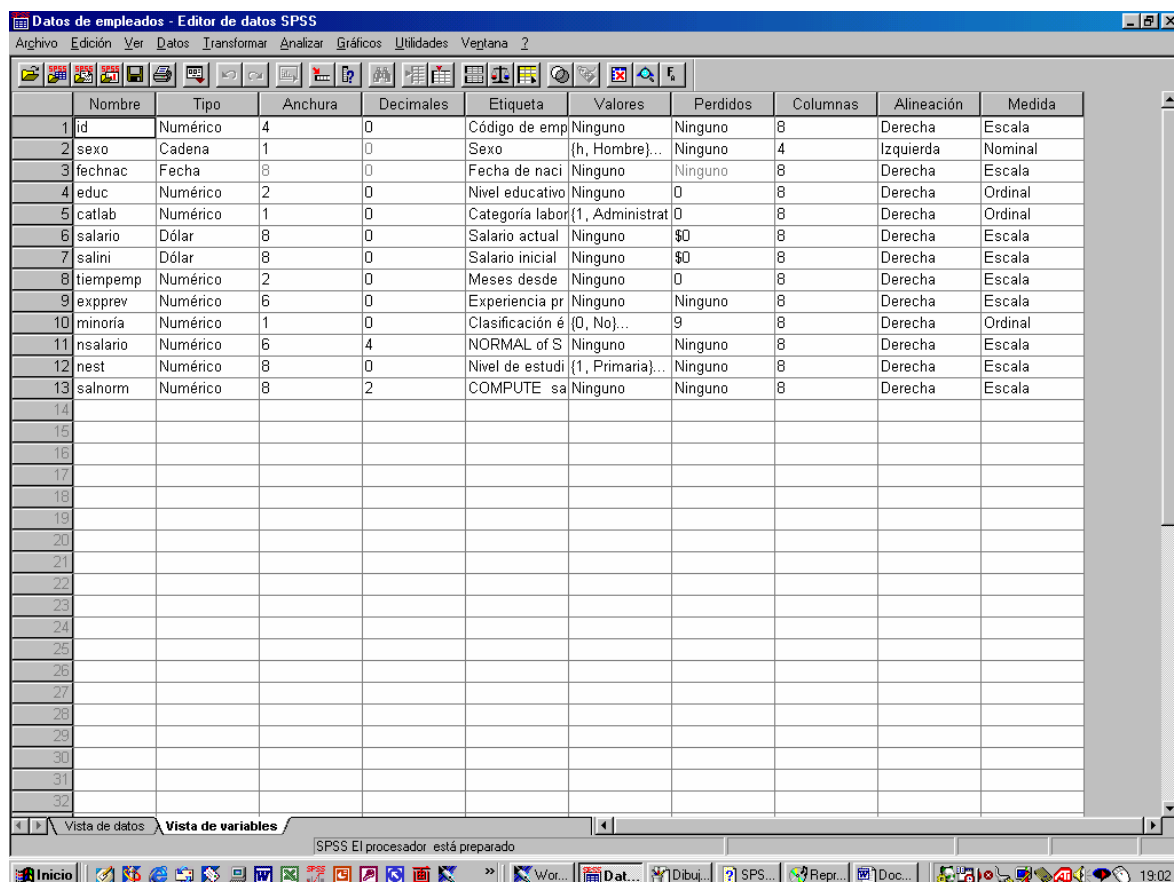


Figura 1.4. Ventana de definición de variables en el Editor de Datos

### 1.2 Definición de Variables


La definición de variables se efectúa en la ventana correspondiente a la Vista de variables en el Editor de datos. A continuación se dan una serie de directrices.

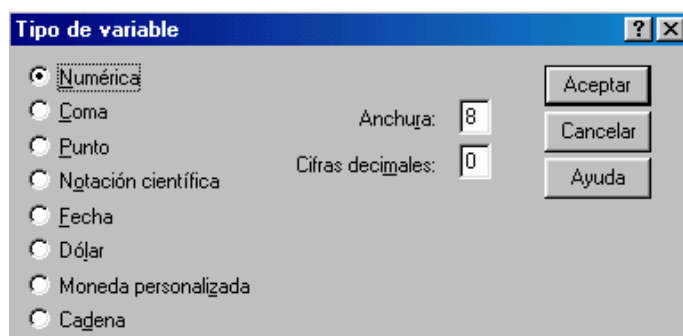
Para los nombres de variable se aplican las siguientes normas:

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

Respecto al Tipo de Variable se pueden elegir entre 8 tipos diferentes:

- **Numérico.** Una variable cuyos valores son números. Los valores se muestran en formato numérico estándar. El Editor de datos acepta valores numéricos en formato estándar o en notación científica.
- **Coma.** Una variable numérica cuyos valores se muestran con comas que delimitan cada tres posiciones y con el punto como delimitador decimal. El Editor de datos acepta valores numéricos para este tipo de variables con o sin comas, o bien en notación científica.
- **Punto.** Una variable numérica cuyos valores se muestran con puntos que delimitan cada tres posiciones y con la coma como delimitador decimal. El Editor de datos acepta valores numéricos para este tipo de variables con o sin puntos, o bien en notación científica.
- **Notación científica.** Una variable numérica cuyos valores se muestran con una E 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. Por ejemplo, 123, 1,23E2, 1,23D2, 1,23E+2 e incluso 1,23+2.
- **Fecha.** Una variable numérica cuyos valores se muestran en uno de los diferentes formatos de fecha\_calendario y hora\_reloj. Seleccione un formato de la lista. Puede introducir las fechas utilizando como delimitadores: barras, guiones, puntos, comas o espacios. El rango de siglo para los valores de año de dos dígitos está determinado por la configuración de las Opciones (menú Edición, Opciones, pestaña Datos).
- **Moneda personalizada.** Una variable numérica cuyos valores se muestran en uno de los formatos de moneda personalizados que se hayan definido previamente en la pestaña Moneda del cuadro de diálogo Opciones. Los caracteres definidos en la moneda personalizada no se pueden emplear en la introducción de datos pero sí se mostrarán en el Editor de Datos.
- **Cadena.** Variable cuyos valores no son numéricos y, por ello, no se utilizan en los cálculos. Pueden contener cualquier carácter siempre que no se exceda la longitud definida. Las mayúsculas y la minúsculas se consideran diferentes. También son conocidas como variables alfanuméricas.

Para definir el tipo se pulsa en la celda de intersección entre la variable y la columna, y una vez señalada la celda se pulsa en el icono que se muestra a la derecha . Al pulsar este icono se muestra el cuadro con todos los tipos de variables como el que se muestra en la Figura 1.5.

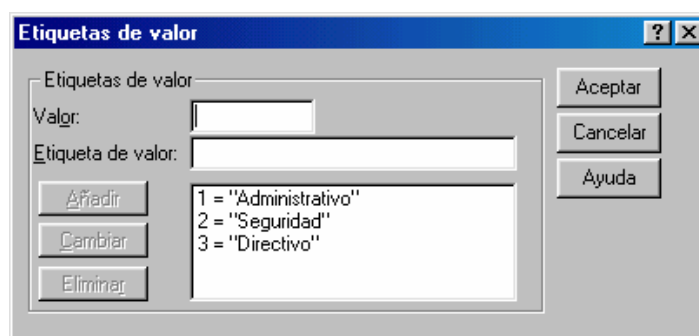


**Figura 1.5 Cuadro de definición del tipo de variable**

Las columnas designadas como **Anchura** y **Decimales**, se emplean para especificar la anchura y el número de decimales que contiene en las variables de tipo *Numérico*, *Coma*, *Punto*, *Notación científica*, *Dólar* y *Moneda personalizada*. Para las variables del tipo *Fecha*, se puede elegir entre un amplio abanico de formatos, y para las variables de tipo *Cadena* únicamente hay que especificar el número de caracteres máximos que tendrá dicha variable.

En la columna **Etiqueta**, se puede escribir un nombre para cada variable más descriptivo que el que proporcionan los 8 caracteres máximos del nombre de la variable.

En la columna **Valores**, se puede dar nombre a los valores numéricos de las variables nominales u ordinales. En el archivo *Datos de empleados* hay una serie de variables que son nominales (o categóricas), como por ejemplo *sexo*, *catlab* o *minoría*. Estas variables se han codificado numéricamente, pero los números asignados no tienen propiedades matemáticas, sino que representan categorías de las variables. Así *catlab* (categoría laboral), se ha codificado como 1, 2 ó 3, según el sujeto sea *Administrativo*, de *Seguridad*, o *Directivo*, respectivamente. Para asignar etiquetas a los valores, se pulsa en la celda correspondiente a la variable y se accede al cuadro que se muestra en la Figura 1.6.



**Figura 1.6. Cuadro para etiquetar los valores de variables nominales u ordinales**

En muchas ocasiones no siempre se puede registrar para todas las variables todas las respuestas de los sujetos, bien porque el valor no se haya registrado o bien porque el sujeto se haya negado a contestar a alguna cuestión; estos casos no tienen validez de cara a los análisis y es preciso identificarlos de alguna manera. Una



opción que permite incluso identificar el origen de estos casos (si el registro se ha perdido, si el sujeto no sabe o no contesta, etc.) es la columna designada como **Perdidos**. Al pulsar en la celda correspondiente de la variable con dicha columna se activa a la derecha el mismo icono con los puntos suspensivos que al pulsar con el ratón nos lleva al cuadro que se muestra en la Figura 1.7.

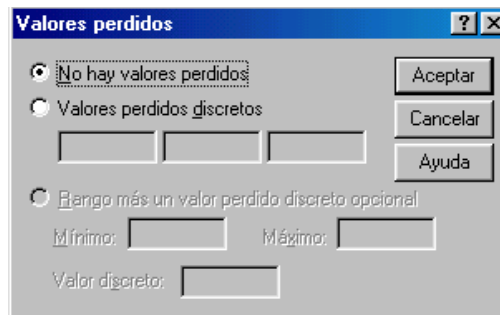


Figura 1.7 Cuadro para definir los valores perdidos

Se observa en este cuadro que se puede especificar como perdidos varios valores discretos, un rango de valores o un solo valor. El analista en cada caso determinará cuál de las opciones es más adecuada.

Por último, la columna **Alineación** permite definir en que posición de la celda (derecha, centro, izquierda) se visualiza el dato en el Editor de Datos. Y la columna **Escala**, permite determinar cómo es la variable: de escala (intervalo o razón), ordinal o nominal.

### 1.3 Definición y ejecución de un procedimiento

Para definir cualquier procedimiento de análisis estadístico, lo primero es disponer de datos en el Editor y, a continuación, elegir un procedimiento estadístico en la opción correspondiente del menú principal. En esta primera sesión confeccionaremos una distribución de frecuencias de la variable Categoría Laboral, del archivo *Datos de empleados*, para ello se sigue la secuencia:

#### **Analizar – Estadísticos descriptivos – Frecuencias**

y se muestra el cuadro de diálogo de la Figura 1.8.

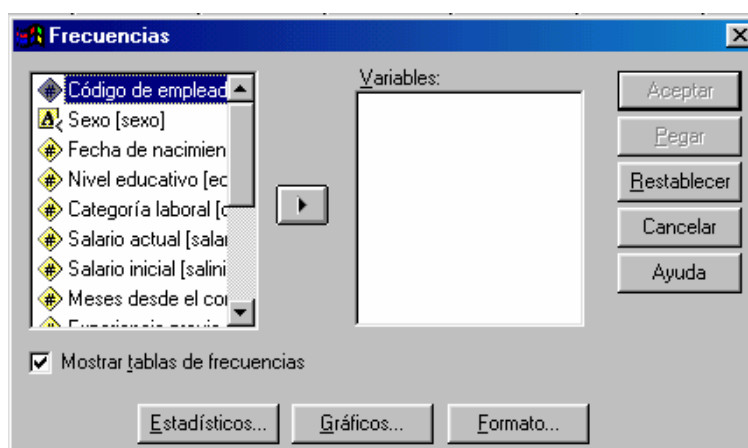


Fig. 1.8 Cuadro de diálogo del procedimiento Frecuencias

En la ventana de la izquierda se muestra la lista de variables que contiene el archivo de trabajo, de entre las cuales seleccionaremos la/s que se quiere/n analizar. Para realizar la selección, se marca cada variable con el puntero del ratón, y se traslada a la lista Variables, mediante la flecha intermedia. Cuando se han pasado las variables a la lista de variables se puede especificar los estadísticos descriptivos y los gráficos que se deseen, pulsando los botones correspondientes en la parte inferior del cuadro de diálogo. Los cuadros a los que se accede son los que se muestran en la Figura 1.9.

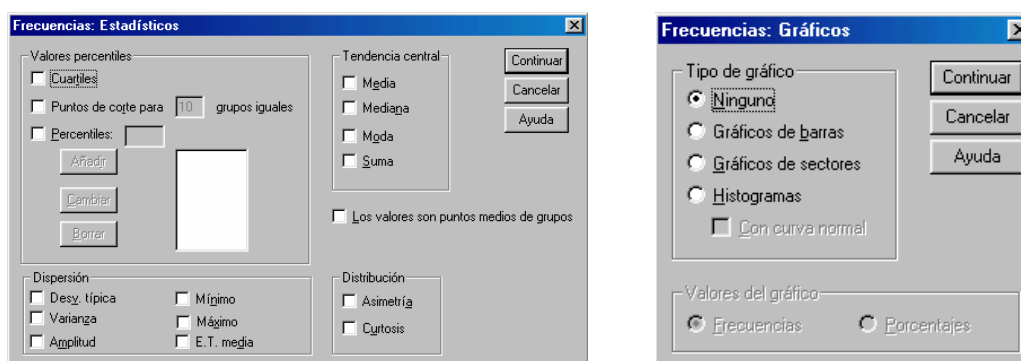


Fig. 1.9 Cuadros de diálogo de estadísticos (izquierda) y de gráficos del procedimiento Frecuencias

En el cuadro de estadísticos podemos señalar cualquiera de los que cuantifican los cuatro aspectos básicos de las distribuciones: los de posición (percentiles), los de tendencia central, los de variabilidad o dispersión y los de forma de la distribución (asimetría y curtosis). Como se ve en la Figura 1.9, por defecto no hay señalado ningún estadístico, y dado que la variable es categórica, tampoco lo vamos a requerir

Respecto a las opciones de gráficos, se puede elegir entre tres tipos, según sea el nivel de medida de la variable. Por defecto, la opción es no confeccionar ningún gráfico.

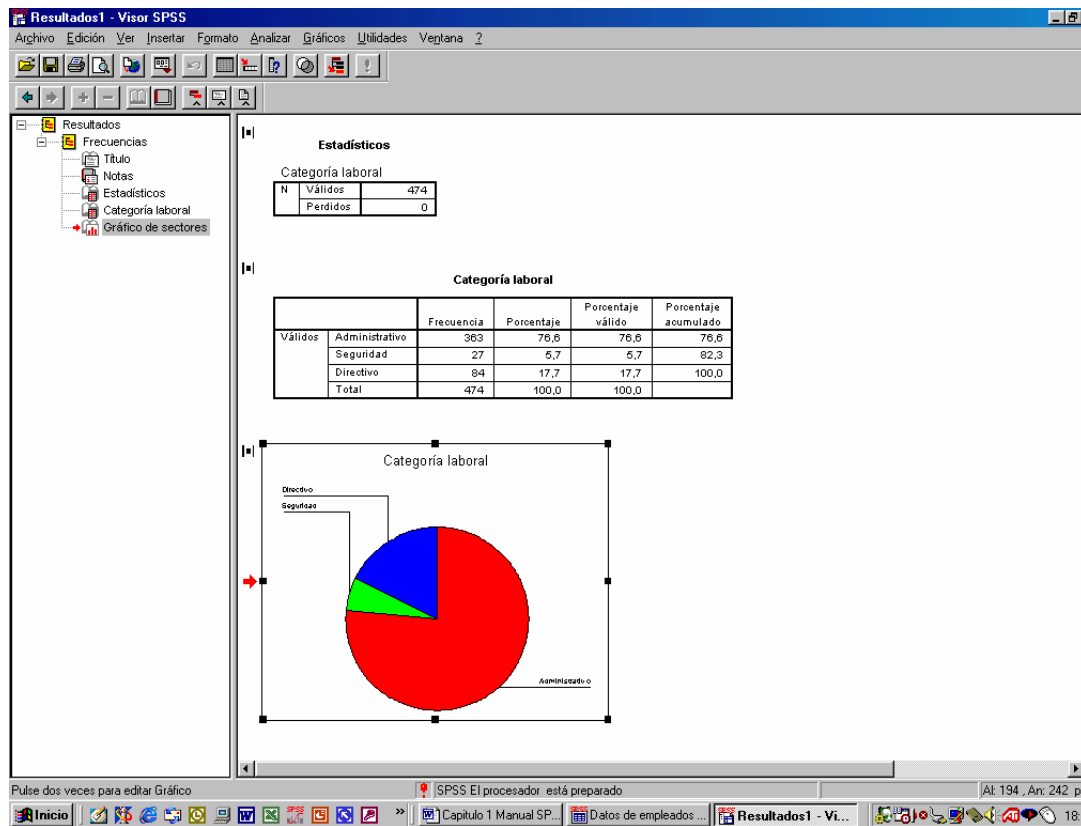
Pulsando, por último, el botón **Formato**, se puede elegir entre varios criterios de ordenación de la tabla de distribución, e incluso optar por no confeccionar

distribución alguna. Por defecto la opción es la de ordenación ascendente de valores

Una vez que está seleccionada la variable y señaladas todas las opciones, de estadísticos, de gráficos y de formato, pulsamos el botón **Aceptar** de la ventana de Frecuencias (Figura 1.7) y entramos en el interface de SPSS, denominado *Visor de SPSS*, cuya facilidad operativa es una de las varias características que lo distinguen favorablemente de las versiones 6 y anteriores.

## 1.4 Navegando por los Resultados

Como se ha dicho, cuando se pulsa el botón **Aceptar**, después de haber configurado las opciones del procedimiento requerido (en esta primera sesión una simple distribución de frecuencias, con su gráfico de pastel), el resultado se muestra en el *Visor*, cuyo aspecto se muestra en la Figura 1.10. La variable seleccionada para analizar es *Categoría laboral*, del archivo *Datos de empleados*.



**Fig. 1.10 Visor de SPSS con algunos resultados del procedimiento Frecuencia**

Este interface consta de dos ventanas: la de la izquierda, con estructura de árbol, es, digamos, el guión o índice de los resultados que se muestran en la pantalla de la derecha. En el índice podemos señalar con el ratón cualquiera de los apartados, y verlo recuadrado en la ventana de la derecha. En la figura se puede ver señalado **Gráficos de sectores**, y en la de la derecha, el diagrama de sectores recuadrado y con una flecha de señal a la izquierda del recuadro. De este modo podemos navegar por los resultados con un simple clic del ratón en la parte que nos interese en cada momento.

### 1.5 Terminar una sesión con SPSS

Cuando ya se han cumplido los objetivos del análisis que hayamos podido efectuar con SPSS y se va a salir del programa, es conveniente guardar el trabajo realizado. Como ya se ha visto son varios los ámbitos en los que nos movemos en las sesiones de análisis, aunque sólo hemos visto dos de ellos: por un lado, el *Editor de datos*, y por otro, los resultados de los análisis que se muestran en el *Visor*. En el *Editor* se muestran los datos que hayamos leído, caso de que estuvieran almacenados en un archivo, o que hayamos escrito en el propio *Editor*. Respecto de los datos, sólo interesa archivarlos de nuevo cuando se ha efectuado alguna modificación de los mismos (recodificación de variables, creación de nuevas variables, etc.); respecto de los resultados, el usuario determinará en cada momento si es conveniente su archivo para una posterior utilización.