



4

Introducción al paquete estadístico SPSS

Neus Canal Díaz

4.1. Introducción

El objetivo de este capítulo es realizar una breve introducción al programa SPSS para facilitar así al lector su utilización básica así como el seguimiento de los posteriores capítulos basados en este paquete estadístico. Vamos a describir opciones básicas del programa como la introducción de datos y la captura desde otros programas o hojas de cálculo, así como otras operaciones básicas pero imprescindibles en el desarrollo de un análisis estadístico, tales como la creación de variables, la recodificación de variables, la ordenación de los datos, la fusión de archivos, etc. Por último, se expondrá la forma de realizar un análisis estadístico primario o básico, sin profundizar en análisis multivariante, que se verá en más detalle en los capítulos dedicados a cada tema. Todo ello lo veremos a partir de los menús de los que dispone el programa, por ser la opción más didáctica y práctica para los usuarios con conocimientos básicos en informática y estadística.

4.2. Ventanas y barras de menús

SPSS trabaja con 3 tipos de ventanas distintas, en cada una se dispone de un

65





menú principal con distintas opciones (según ventana), la barra de herramientas con una serie de iconos para acceder a algunas tareas de forma directa y la barra de estado que indica el estado del procesador. Todas las ventanas tienen un menú de ayuda (?) en la barra de menús. *Temas* (Topics) proporciona acceso a las pestañas Contenido, Índice y Buscar. Tutorial proporciona acceso al tutorial de introducción.

4.2.1. Ventana de editor de datos (.sav)

El editor de datos (data editor) es la ventana del SPSS que contiene el archivo de datos en que se basan todos los análisis. Dispone de un sistema de visualización tipo hoja de cálculo diseñada para crear y editar archivos de datos. Se abre automáticamente al iniciar la sesión.

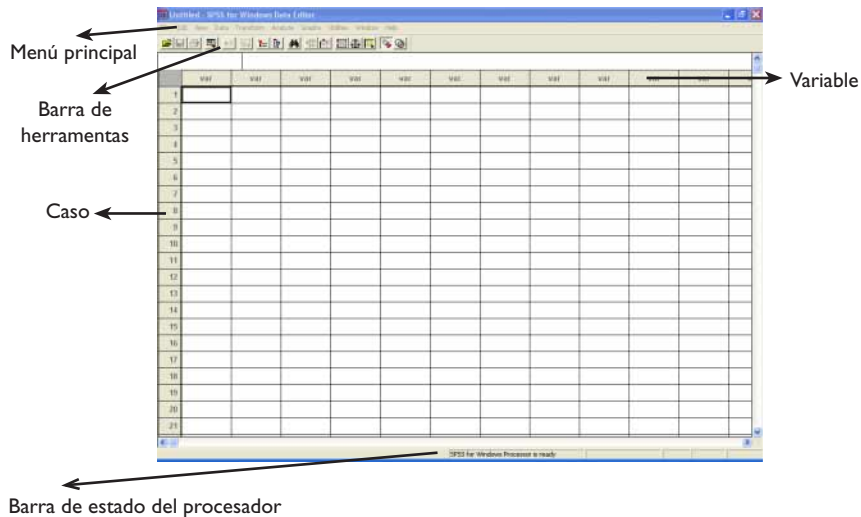


Figura 2. Ventana del editor de datos

El editor de datos se estructura de manera que:

- Las filas representan casos, por ejemplo sujetos
- Las columnas representan variables, por ejemplo cada pregunta de un cuestionario
- Cada casilla contiene un valor individual que corresponde a un determinado caso de una determinada variable. Las casillas solo pueden contener valores o caracteres correspondientes a variables numéricas o alfanuméricas (no fórmulas). Las casillas vacías de una variable numérica se consideran valores perdidos (missings de sistema).





4.2.1.1. Menú principal

Veamos los principales subapartados dentro del menú principal:

- **Archivo (File):** se puede abrir (open), guardar (save as) o imprimir (print) archivos como en cualquier otro programa. Si lo que se desea es abrir un archivo de datos ya existente se pulsa ARCHIVO>ABRIR (FILE>OPEN). Para crear un archivo nuevo ARCHIVO>NUEVO>DATOS (FILE>NEW>DATA). Para importar datos externos, como pueden ser de un archivo de hoja de cálculo Excel debe pulsar ARCHIVO>ABRIR y seleccionar el tipo .xls. Una vez elegido debe indicarse al SPSS si la primera fila del archivo de Excel contiene el nombre de las variables y el rango a leer en la misma si los datos no comienzan en la primera fila ni en la primera columna. La versión 9 de SPSS solo puede leer las versiones 4 o anteriores, por lo que la hoja de cálculo deberá guardarse en una de estas versiones.

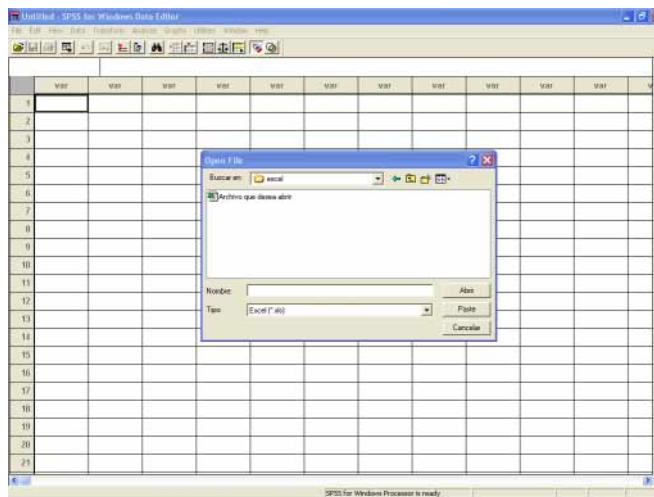


Figura 3. Cuadro de diálogo para abrir archivo

- **Edición (Edit):** se puede copiar (copy), pegar (paste), eliminar (clear) y buscar texto (find). La opción deshacer (undo) permite anular la última operación realizada.
- **Datos (Data):** Vamos a detenernos un poco más en esta opción del menú para describir más detalladamente las subopciones que más pueden interesar al usuario, como puede ser la definición de variables (nombre, etiquetas, formatos) inserción de nuevos casos y variables, unir archivos, ordenar, segmentar y seleccionar casos.



- **Datos>definir variable (Data>Define variable):** para crear y definir una variable tenemos dos opciones válidas. Podemos seleccionar DATOS>DEFINIR VARIABLE (DATA>DEFINE VARIABLE) o pulsar dos veces en la parte superior de la columna (indicado por 'var' si no se ha creado ninguna variable). El cuadro de diálogo que aparece es el siguiente:

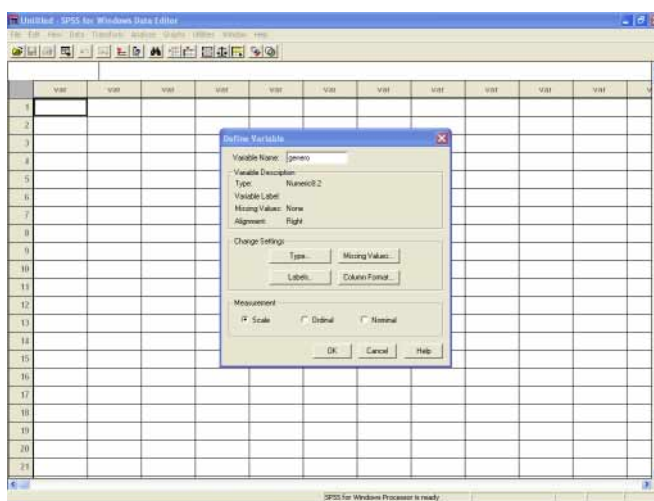


Figura 4. Cuadro de diálogo para la creación y definición de una variable

Nombre de la variable (variable name): Puede introducir el nombre de la variable, que deberá empezar por una letra, no acceder de los 8 caracteres y no contener espacios en blanco ni símbolos especiales. SPSS no permite nombres duplicados y no distingue entre minúsculas y mayúsculas, por ejemplo genero, como se muestra en la Figura 4.

Tipo (type): para especificar los tipos de datos de cada variable. Por defecto, SPSS asume que son numéricas (Figura 5). Como *numérico* (numeric) acepta cualquier número, el símbolo +, el símbolo – y el separador decimal (punto o coma, dependiendo de las especificaciones establecidas en Windows). En anchura (width) se puede especificar el total de dígitos y decimales que contendrá como máximo la variable. Como *fecha* (date) admite fechas y horas en distintos formatos y como *cadena* (string) cualquier carácter, especificando en anchura el número máximo de caracteres.



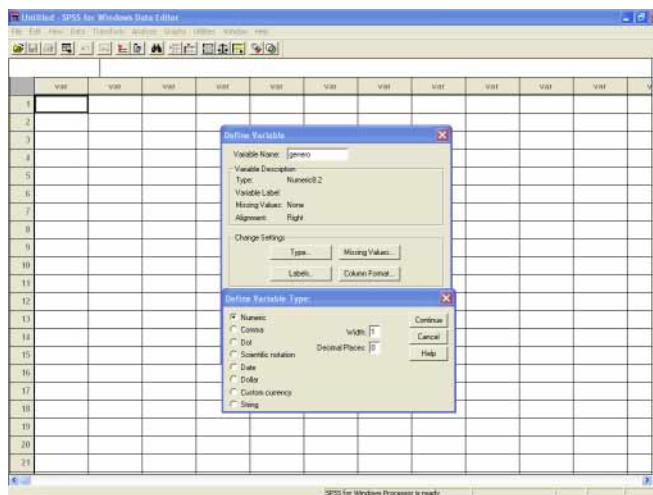


Figura 5. Cuadro de diálogo para tipo de variable

Etiquetas (labels): Pueden asignarse *etiquetas descriptivas a las variables* (variable label), dado que en muchos casos el nombre de una variable es insuficiente para reconocer el significado, o *etiquetas de valor* (value label). Por ejemplo: variable de género (1 hombre, 2 mujer). Entonces introduciríamos en *etiqueta de variable* 'Género del paciente' (para ver la etiqueta solo tendrá que poner el cursor sobre el nombre de la variable una vez creada), en *valor* (value) el 1 y en *etiqueta de valor* hombre (para visualizar las etiquetas en el archivo de datos tendrá que activar el icono de valor de etiquetas, segundo por la derecha del archivo de datos). Al pulsar en añadir (add), la etiqueta introducida pasa al recuadro de abajo, pudiendo entonces introducir otro valor con su etiqueta correspondiente, valor 2 etiqueta de valor 'mujer' (Figura 6).

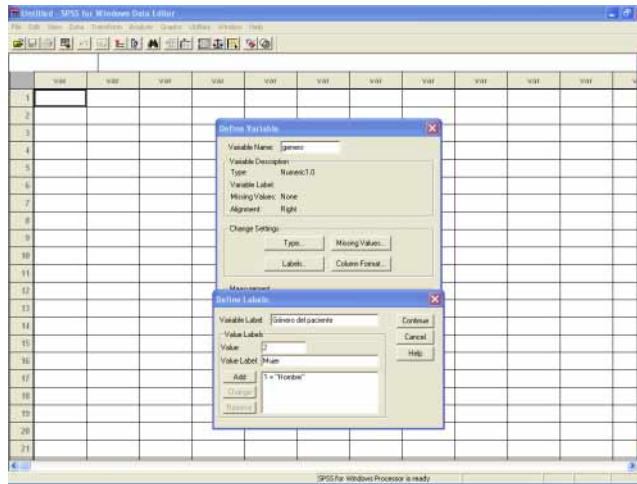


Figura 6. Cuadro de diálogo para definir etiquetas de valor o de variable

Valores perdidos (missing values): se refiere a valores perdidos definidos por el usuario. En algunos casos puede resultar útil distinguir entre diferentes tipos de valores perdidos. Por ejemplo: en el caso de un cuestionario podrían ser los sujetos que no conocen la respuesta (9), los que no responden (perdido de sistema) y los que no desean responder (99). Se pueden definir 9 y 99 como valores perdidos del usuario siempre y cuando éstos no sean valores válidos de la variable (Figura 7).

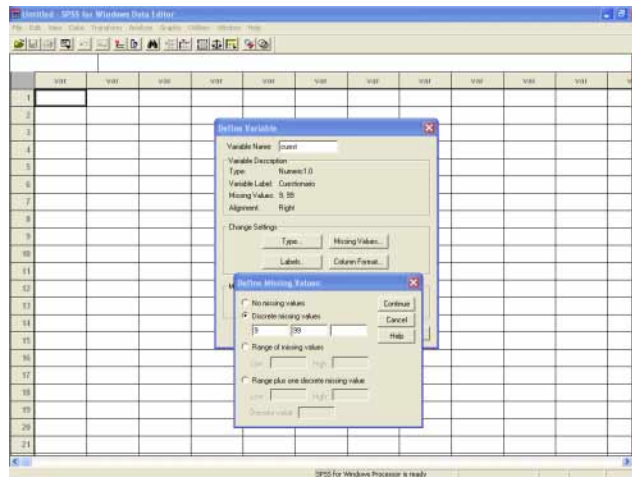


Figura 7. Cuadro de diálogo para la definición de valores perdidos de usuario



– **Datos>insertar caso (Data>Insert Case):** la introducción de datos en una casilla de una fila vacía crea automáticamente un nuevo caso. El programa inserta el valor perdido por el sistema en todas las demás variables. También puede insertarse un nuevo caso entre casos existentes con DATOS>INSERTAR CASO (DATA>INSERT CASE).

– **Datos>insertar nueva variable (Data>Insert variable):** al introducir un dato en una casilla de una columna vacía crea automáticamente por defecto una nueva variable, insertando un valor perdido por sistema en todos los demás casos de la nueva variable. También se puede insertar con DATOS>INSERTAR VARIABLE (DATA>INSERT VARIABLE). El nombre y formato de la variable se modifican pulsando 2 veces sobre el nombre de la variable (como se ha comentado en definir variable), que por defecto será var00001, var00002,... y de tipo numérico.

– **Datos>ordenar casos (Data>Sort Cases):** Ordena los casos (filas) del archivo de datos basándose en los valores de una o más variables (Figura 8).

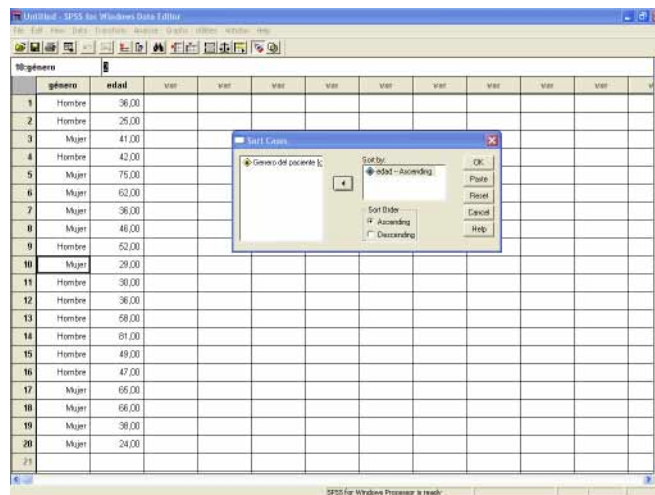


Figura 8. Cuadro de diálogo para ordenar casos

– **Datos>fundir archivos (Data>Merge files):** Hay dos maneras diferentes de unir o juntar datos de dos archivos, fundir archivos que tengan las mismas variables pero casos diferentes, y fundir archivos que contengan los mismos casos pero variables diferentes.

Para *añadir casos* (Add Cases), un ejemplo podría ser disponer de un archivo de datos de pacientes para cada hospital, los dos archivos de trabajo deben



tener las mismas variables y mismo formato, a no ser así, el SPSS indica cuales son las variables desemparejadas, indicando mediante un asterisco (*) las variables del archivo de trabajo de datos de SPSS que está abierta (por lo tanto visualizada) y con un signo mas (+) las variables del archivo externo que queremos unir, como se muestra en la siguiente figura.

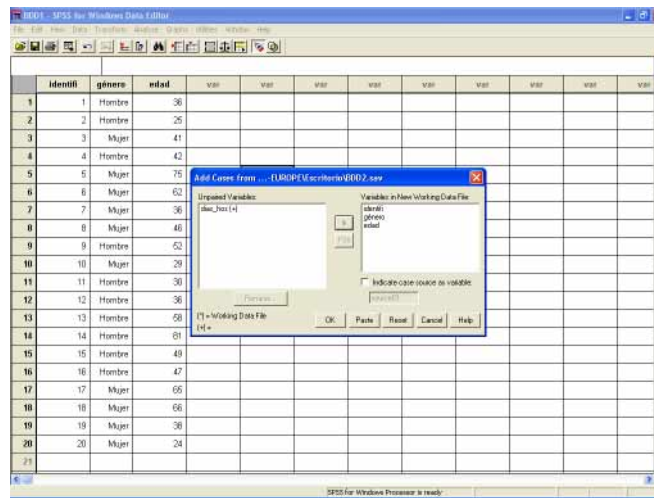


Figura 9. Cuadro de diálogo para añadir casos desde un archivo externo

Para *añadir variables* (add Variables), por ejemplo un archivo de trabajo para cada visita de los pacientes, debemos tener una o más variables que identifiquen el caso en ambos archivos (como podrían ser número de hospital o número de paciente) y pasarlas a variables clave una vez seleccionada la opción *emparejar los casos* (Match cases on key variables in sorted files). El origen de las variables se indica como en el caso anterior (* o +). **Es muy importante que ambos archivos estén ordenados por orden ascendente mediante las variables clave y no existan casos repetidos según las variables clave en la base de datos utilizada como archivo de trabajo** (en el ejemplo de la Figura 10 'identifi' sería una variable clave).



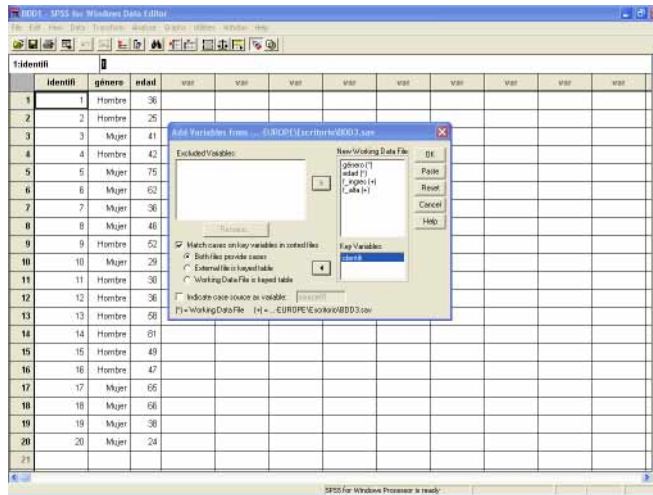
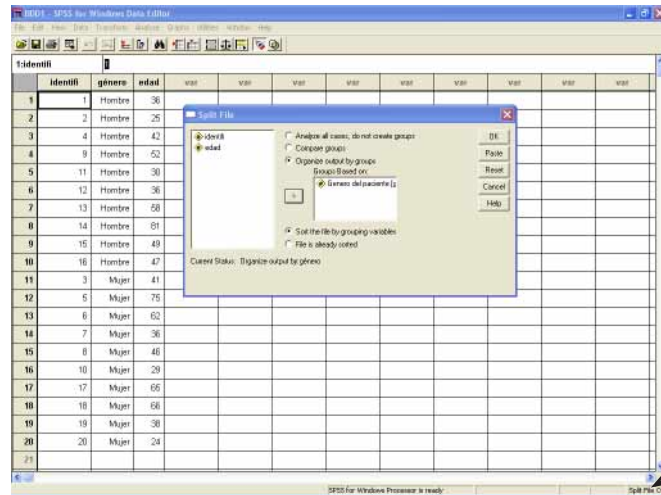


Figura 10. Cuadro de diálogo para añadir variables desde un archivo externo

– **Datos>segmentar archivo (Data>Split file):** esta opción divide el archivo de datos en distintos grupos para el análisis, basándose en una o más variables de agrupación, por ejemplo obtener la edad media de los hombres y las mujeres de manera separada. El archivo de datos debe estar ordenado por las variables de agrupación, si no lo está, SPSS nos da esta opción dentro del cuadro de diálogo que aparece al pulsar **DATOS>SEGMENTAR ARCHIVO (DATA>SLITE FILE)** (Figura 11). Si escogemos *Comparar los grupos* (compare groups) los resultados se presentaran en una sola tabla, mientras que si escogemos *Organizar los resultados por grupos* (Organiza output by groups) los resultados nos aparecerán en tablas diferentes por cada categoría de la variable de agrupación. Una vez seleccionada una de estas opciones, debe pasarse la variable de agrupación al cuadro *Grupos basados en* (groups based on) y aceptar (OK). En la parte inferior derecha del editor de datos se nos indica la segmentación del archivo. Para desactivar esta opción se pulsa **DATOS>SEGMENTAR ARCHIVO (DATA>SPLITE FILE)** y escogemos *Analizar todos los casos* (Analyze all cases).



Segmentación del archivo

Figura 11. Cuadro de diálogo para la segmentación de un archivo

– **Datos>seleccionar casos (Data>Select Cases):** hay varios métodos para seleccionar un grupo de casos. El más común es a partir de una expresión condicional, *si se satisface la condición* (if condition is satisfied). Si el resultado de una expresión condicional es verdadero, el caso se selecciona, si es falso el caso no se selecciona. La expresión condicional utiliza alguno de los operadores de relación (<, >, <=, >=, = y <>), pudiendo incluir también nombre de variable, constantes o funciones numéricas. Un ejemplo de expresión condicional sería $sexo=1$, donde 1 significa hombre, por lo que sólo se seleccionarán los hombres (Figura 12). Esta selección puede ser un filtro temporal (opción *filtrados* (filtered)), de manera que los casos no seleccionados permanecerán en el archivo de datos pero quedarán marcados con una raya sobre el número de caso o fila y no se incluirán en los posteriores análisis, o de manera permanente (opción *eliminados* (deleted)) y desaparecerán (no recuperables) del archivo de datos. En el caso de ser un filtro temporal se genera de forma sistemática una variable que permite diferenciar los casos seleccionados de los no seleccionados. Para desactivar una selección de filtro se debe pulsar **DATOS>SELECCIONAR CASOS (DATA>SELECT CASES)** y la subopción **Todos los casos (All cases)**.



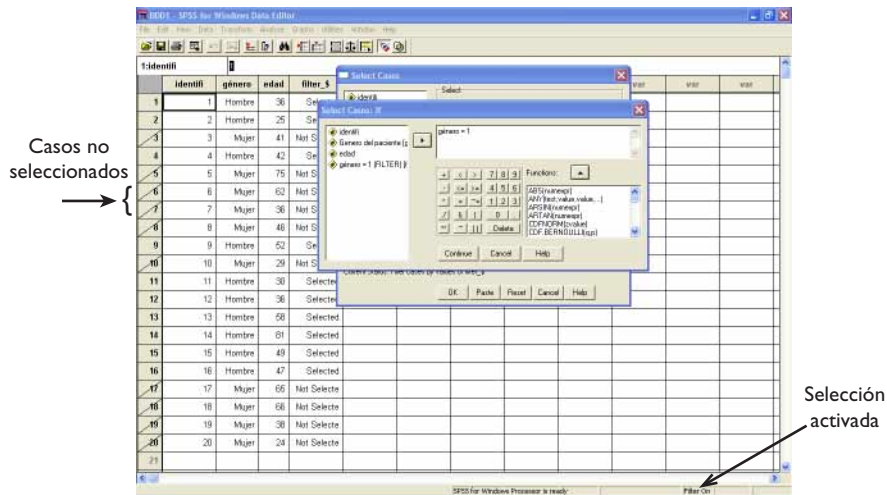


Figura 12. Cuadro de diálogo para la selección de casos

– **Transformar (Transform):** En muchas ocasiones también nos interesará calcular nuevas variables a partir de datos ya almacenados en el archivo de datos o recodificar variables.

– **Transformar>calcular (Transform>compute):** nos permite calcular nuevas variables. Primero debe darse un nombre a la nueva variable en *Variable de destino* (target variable) y especificar el tipo y etiqueta (type&label) (opcional). Debe generarse una expresión numérica para calcular la nueva variable que contendrá variables del archivo de datos, operadores simples y/o funciones predefinidas por SPSS (con el botón derecho del ratón sobre una función el SPSS nos da una pequeña descripción de lo que realiza). Por ejemplo, si deseamos calcular la edad de los pacientes a partir de la fecha de visita del paciente y la fecha de nacimiento, la expresión numérica sería $CTIME.DAYS$ (nombre variable fecha de visita – nombre variable fecha de nacimiento)/365, donde $CTIME.DAYS$ es una función determinada del programa que calcula los días que han pasado entre dos fechas (Figura 13).

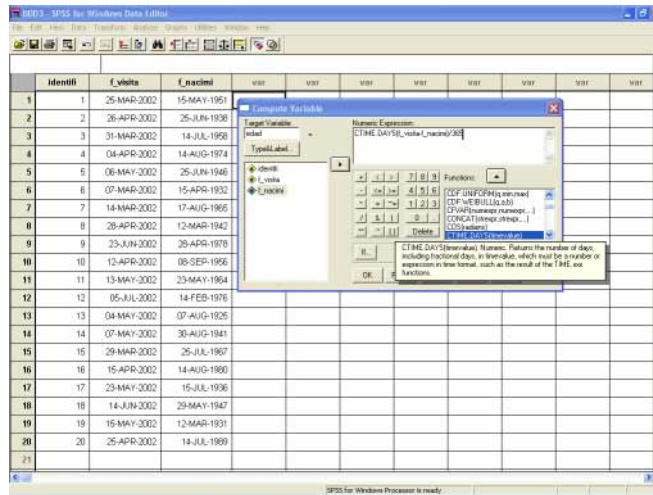


Figura 13. Cuadro de diálogo para el cálculo de nuevas variables a partir de variables existentes

El cuadro de diálogo Si... (If...) nos permitirá aplicar transformaciones de los datos en subconjuntos de casos seleccionados utilizando expresiones condicionales.

– Transformar>recodificar (Transform>Recode): la recodificación es útil para agrupar y combinar categorías. La recodificación puede ser en la misma variable que se desea recodificar o en una variable distinta (una variable nueva). La recodificación en la *misma variable* (Into same variables) reasigna los valores de las variables existentes o agrupa rangos de valores existentes en nuevos valores. Por ejemplo, podría agruparse la edad en rangos de edad (niños, adultos, ancianos,...). Primero debe seleccionarse la variable que se desea recodificar en la ventana de la izquierda. Pulse en valores antiguos y nuevos (Old and New values) y especifique como deben recodificarse los valores. Una vez seleccionados el/los valor/es antiguos y nuevo no debe olvidarse pulsar en añadir (add) (Figura 14). Si se desea puede definirse un subconjunto de casos para recodificarlos (Si... (If...)).

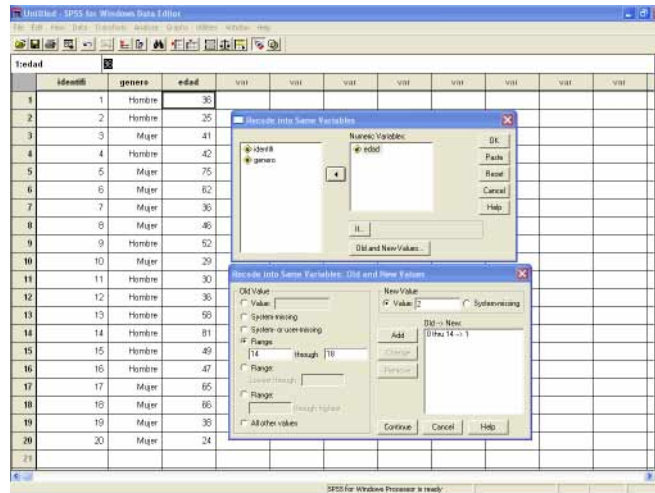


Figura 14. Cuadro de diálogo para la recodificación de valores de una variable

La recodificación en *distintas variables* (Into different variables) asigna los valores de variables existentes o agrupa rangos de variables existentes a una nueva variable. Por ejemplo, agrupar la edad en una nueva variable que contenga categorías de rangos de edad. Los pasos son los mismos utilizados en el caso anterior, solo hace falta indicar el nombre y etiqueta de la nueva variable.

– **Analizar (Analyze):** en esta opción se encuentran la mayoría de procedimientos estadísticos, mucho de ellos los veremos en los capítulos del libro dedicados a ello.

– **Analizar>estadísticos descriptivos (Analyze>Descriptive Statistics):** Habitualmente el primer tipo de análisis que se realiza sobre las variables objeto de estudio es el cálculo de estadísticos descriptivos (media, mediana, moda, desviación típica, ...) en el caso de variables continuas o cuantitativas, y frecuencias en el caso de variables categóricas. Para iniciar el cálculo de los estadísticos descriptivos básicos de las variables cuantitativas, puede utilizarse la opción ANALIZAR>ESTADÍSTICOS DESCRIPTIVOS>FRECUENCIAS (ANALYZE>DESCRIPTIVE STATISTICS >FREQUENCIES) o bien la alternativa ANALIZAR>ESTADÍSTICOS DESCRIPTIVOS>DESCRIPTIVOS (ANALYZE>DESCRIPTIVE STATISTICS>DESCRIPTIVES). La primera opción permite calcular un mayor número de estadísticos. En la primera de las pantallas debe seleccionarse el conjunto de variables para las que se desea elaborar los estadísticos (si aceptamos en este paso obtendríamos la tabla de frecuencias para las variables categóricas seleccionadas). La subopción *estadísticos* (Statistics) contiene





ne el menú de estadísticos disponibles, donde debemos seleccionar aquellos que deseamos calcular. Pulsando el botón de gráficos (Charts) podremos crear distintos tipos de gráficos que nos ayudarán a interpretar los resultados. Por ejemplo, el histograma nos permitirá observar gráficamente la distribución de la variable para valorar su Normalidad.

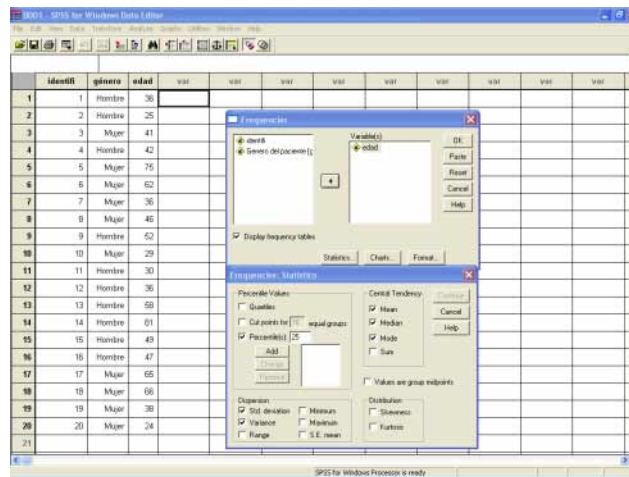


Figura 15. Cuadro de diálogo para el cálculo de estadísticos básicos

-Gráficos (Graphs): Esta opción permite escoger entre distintos tipos de diagramas, como son diagramas de barras, líneas, áreas o sectores.

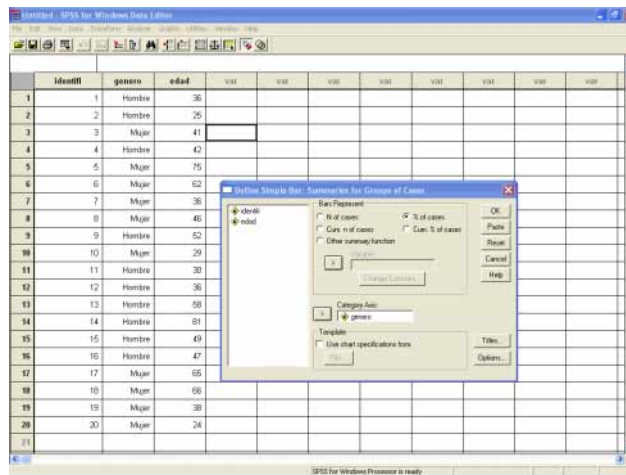


Figura 16. Cuadro de diálogo para realizar un gráfico de barras





4.2.1.2. Introducción de datos

Para introducir directamente los datos en SPSS es siempre mejor definir a priori las propiedades de las variables. Para cada variable deben definirse los siguientes aspectos:

- 1- Asignar un nombre y una etiqueta a la variable
- 2- Identificar el tipo de variable
- 3- Definir sus posibles valores y etiquetas
- 4- Definir el o los valores perdidos

Para introducir o modificar un dato colocamos el cursor sobre la celda implicada y solo tenemos que escribir su valor que se mostrará en el editor de casillas situado en la parte superior del editor de datos, y no se registrará hasta que pulse **intro** o se seleccione otra celda. Para moverse por las celdas se utilizan las flechas de desplazamiento del teclado. La forma habitual de introducir nuevos datos es por variables.

4.2.2. Ventana de sintaxis de comandos (.sps)

Para crear una ventana de sintaxis sólo tiene que pulsar ABRIR>NUEVO>SINTAXIS (FILE>NEW>SINTAX). Desde esta ventana puede acceder a la mayoría de comandos desde los menús y cuadros de diálogo. El lenguaje de comandos permite guardar los trabajos en un archivo de sintaxis, con lo que podrá repetir los análisis en otro momento. Un archivo de sintaxis es simplemente un archivo de texto que contiene comandos. Aunque es posible abrir una ventana de sintaxis y escribir comandos, es más sencillo permitir que el programa le ayude a construir un archivo de sintaxis. La opción más fácil es a partir de la opción *Pegar* (Paste) que aparece en los cuadros de diálogo. Una vez disponemos de la sintaxis de un comando (cada comando debe terminar con un punto '.'), se selecciona y se ejecuta mediante el icono indicado en la Figura 16.





Icono ejecución ←

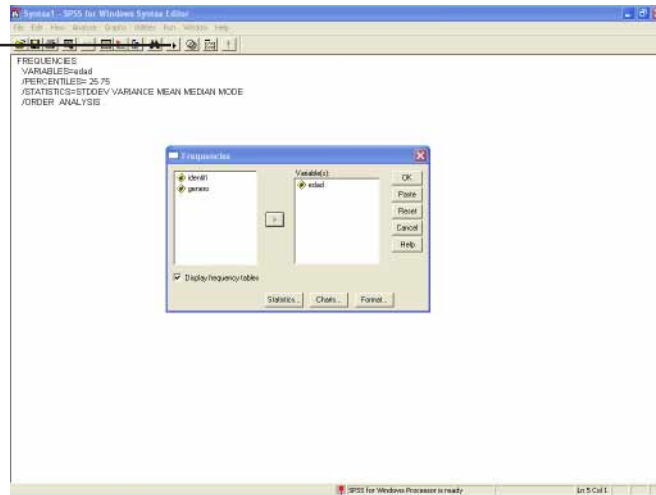


Figura 17. Ventana de sintaxis con el comando del cuadro de diálogo para el cálculo de estadísticos básicos (opción pegar/paste)

4.2.3. Ventana de visor de resultados (.spo)

Cuando se ejecuta un procedimiento (como el utilizado anteriormente para el cálculo de estadísticos descriptivos), los resultados se muestran en una ventana llamada visor de resultados. Desde esta ventana también puede acceder a la mayoría de comandos desde los menús y cuadros de diálogo. En la ventana de resultados se puede crear un documento que contenga los resultados que desee, organizarlos a su medida y guardarlo.

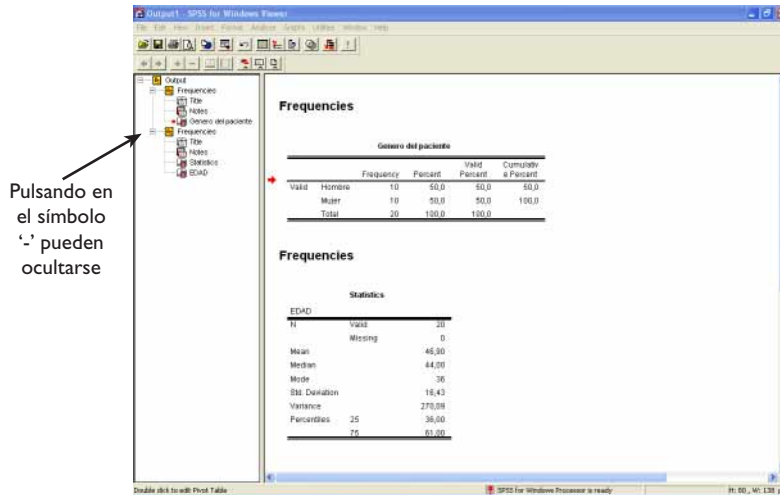


Figura 18. Ventana de visor de resultados

4.3. Consideraciones importantes

Todos los procedimientos comentados en este capítulo hacen referencia a la versión 9 del SPSS. La diferencia básica con versiones posteriores es la definición de variables, puesto que el editor de datos presenta en la parte inferior izquierda dos pestañas, Vista de datos y vista de variables. En vista de variables se pueden llevar a cabo todas las tareas relacionadas con la definición de una variable.

Este capítulo no pretende ser un manual exhaustivo del programa SPSS. Lo que pretendíamos era comentar las opciones posiblemente más utilizadas para el procesamiento de datos y análisis estadístico. El SPSS presenta muchas más herramientas, algunas se verán en los distintos capítulos de éste libro, pero también pueden consultarse en cualquier manual del programa.

