

Capítulo 6

Modificar archivos de datos

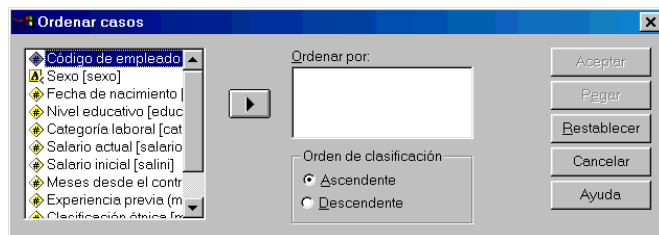
Los archivos de datos no siempre están organizados de forma idónea. En ocasiones podemos desear cambiar el orden de los casos, o transponer las filas y las columnas, o mezclar en uno varios archivos diferentes, o seleccionar sólo unos pocos casos para el análisis, etc. En este capítulo vamos a estudiar un conjunto de acciones relacionadas con el archivo de datos, todas las cuales se encuentran en el menú **Datos**.

Ordenar casos

La opción **Ordenar casos** permite cambiar el orden de los casos (es decir, el orden de las filas del *Editor de datos*) utilizando como criterio una o más variables. Es una opción necesaria para preparar la fusión de archivos (ver más adelante, en este mismo capítulo, el apartado *Fundir archivos*) y puede resultar útil, por ejemplo, si se desea imprimir el archivo de datos con los casos ordenados siguiendo algún criterio de interés. Para ordenar casos:

- ▶ Seleccionar la opción **Ordenar casos...** del menú **Datos** para acceder al cuadro de diálogo *Ordenar casos* que muestra la figura 6.1.

Figura 6.1. Cuadro de diálogo *Ordenar casos*.



Ordenar por. Los casos del archivo se ordenan utilizando como criterio la(s) variable(s) trasladada(s) a esta lista desde lista de variables. Si se utiliza más de una variable, el orden resultante viene determinado por la secuencia de las variables seleccionadas.

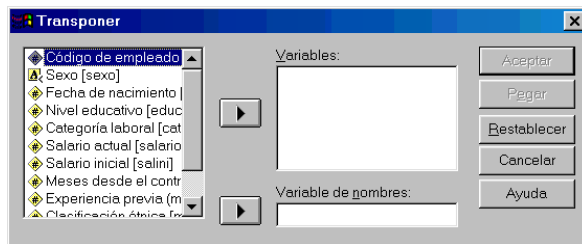
Orden de clasificación. La ordenación puede hacerse de dos maneras. Si el criterio de ordenación es una variable numérica, la opción **Ascendente** ordena los casos desde el más pequeño al mayor. Si el criterio es una variable de cadena, esta opción ordena los casos alfabéticamente (las mayúsculas preceden a las minúsculas). La opción **Descendente** realiza la ordenación de forma inversa a como lo hace la opción **Ascendente**.

Transponer archivos

En los archivos de datos SPSS se asume que las filas representan *casos* y las columnas *variables*. Esto es así también en la mayor parte de las bases de datos y hojas de cálculo. Sin embargo, puede ocurrir que nos veamos en la necesidad de importar datos de alguna fuente externa en la que las filas representen variables y las columnas casos. En ese caso, la opción **Transponer** permite crear un nuevo archivo con los datos transpuestos, es decir, con las filas convertidas en columnas y las columnas en filas. Para transponer el archivo de datos:

- ▶ Seleccionar la opción **Transponer...** del menú **Datos** para acceder al cuadro de diálogo *Trasponer* que muestra la figura 6.2.

Figura 6.2. Cuadro de diálogo *Transponer*.



Variables. Sólo pasan a formar parte del nuevo archivo transpuesto las variables trasladadas a esta lista. Cada variable seleccionada en esta lista pasa a ser un caso. Cada caso del archivo de datos pasa a ser una nueva variable.

Variable de nombres. Por defecto, las nuevas variables (una por cada caso) son nombradas *var001*, *var002*, etc. Pero, opcionalmente, los nuevos nombres pueden ser los valores de alguna de las variables del archivo (antes de ser transpuesto). Para ello, la variable cuyos valores serán los nombres de las nuevas variables debe desplazarse a este cuadro.

Si la variable propuesta es numérica, los nuevos nombres comienzan con la letra “V”. Si el valor excede de 8 caracteres, sólo se toman los 8 primeros. Si los valores de la variable propuesta no son *únicos*, a cada valor repetido se le asigna un número secuencial al final. El *Visor de resultados* ofrece un listado con los nombres asignados las nuevas variables.

Fundir archivos

El SPSS permite combinar en un solo archivo los datos de archivos diferentes. Existen dos posibilidades de combinación o fusión de archivos:

- **Añadir casos.** Consiste en combinar archivos que contienen las mismas variables, pero casos diferentes.
- **Añadir variables.** Consiste en combinar archivos que contienen los mismos casos, pero distintas variables.

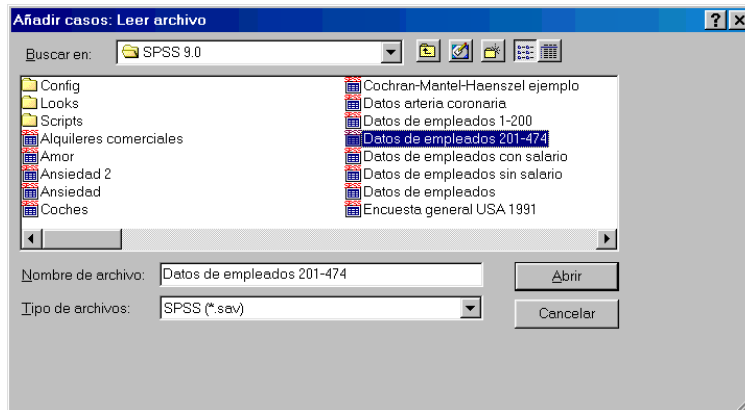
Añadir casos

Nota: Para ilustrar el procedimiento **Fundir archivos > Añadir casos** hemos partido en dos el archivo *Datos de empleados*. Los primeros 200 casos los hemos guardado en el archivo *Datos de empleados 1-200*; y los casos restantes los hemos guardado en el archivo *Datos de empleados 201-474*. Además, la variable *educ* mantiene el nombre en el archivo *Datos de empleados 1-200* pero ha cambiado a *estudios* en el archivo *Datos de empleados 201-474*. Así pues, aunque ambos archivos contienen las mismas variables (si bien una de ellas con nombre diferente), contienen casos distintos.

Para añadir casos debemos comenzar abriendo en el *Editor de datos* uno de los archivos que deseamos combinar. A este primer archivo lo llamaremos **archivo de trabajo**. Utilizaremos como archivo de trabajo *Datos de empleados 1-200*. Tras abrir este archivo, debemos seleccionar el archivo que vamos a fundir con el de trabajo. Para ello:

- Seleccionar la opción **Fundir archivos > Añadir casos...** del menú **Datos** para acceder al cuadro de diálogo *Añadir casos: Leer archivo* que muestra la figura 6.3.

Figura 6.3. Cuadro de diálogo *Añadir casos: Leer archivo*.

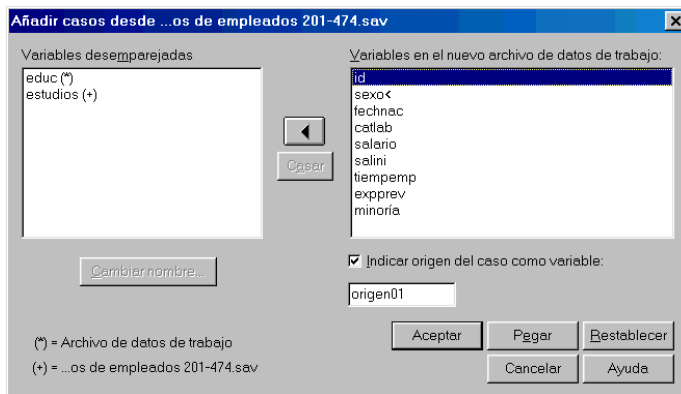


Este cuadro de diálogo contiene un listado de los archivos de datos de la carpeta por defecto. Vemos que en ese listado aparecen los dos archivos que hemos creado a partir del archivo *Datos de empleados*, es decir, los archivos *Datos de empleados 1-200* y *Datos de empleados 201-474*. Desde este cuadro de diálogo es posible seleccionar el archivo de datos cuyos casos deseamos añadir al archivo de trabajo. A este segundo archivo lo llamaremos **archivo externo**. Por supuesto, tanto el archivo de trabajo como el archivo externo deben ser archivos de datos con formato SPSS. Al archivo resultante de la fusión lo llamaremos **archivo combinado**.

- ❑ Seleccionar el archivo *Datos de empleados 201-474* y pulsar el botón **Abrir** para acceder al cuadro de diálogo *Añadir casos desde...* que muestra la figura 6.4.

En el ejemplo de la figura 6.4, el archivo de datos externo seleccionado es *Datos de empleados 201-474*, lo cual queda reflejado en el título del cuadro de diálogo y en la parte inferior izquierda, precedido del signo “+”.

Figura 6.4. Cuadro de diálogo *Añadir casos desde...*



Variables desemparejadas. Contiene un listado de las variables que no serán incluidas en el archivo combinado. Esta lista contiene:

- Variables que se encuentran sólo en uno de los dos archivos (caso del ejemplo).
- Variables definidas como *numéricas* en uno de los archivos y como de *cadena* en el otro: las variables numéricas no pueden combinarse con las de cadena.

- Variables de cadena de distinto ancho. El ancho definido para una variable de cadena debe ser el mismo en ambos archivos.

Las variables del archivo de trabajo aparecen acompañadas de un asterisco (*). Las variables del archivo externo aparecen acompañadas de un signo más (+). Para recordar esto, en la parte inferior izquierda del cuadro de diálogo aparecen los archivos de trabajo y externo precedidos de los símbolos que los identifican.

En el ejemplo de la figura 6.4, la lista **Variables desemparejadas** contiene dos variables: la variable *educ*, que pertenece al archivo de trabajo (*) y la variable *estudios*, que pertenece al archivo externo (+).

Casar. Si, como ocurre en el ejemplo, dos variables distintas contienen la misma información, el SPSS ofrece la posibilidad de emparejar esas dos variables para formar una sola. Para ello, basta con marcar las dos variables que deseamos emparejar y pulsar el botón **Casar** situado debajo del botón flecha. Las dos variables son entonces trasladadas a la lista **Variables en el nuevo archivo de datos de trabajo** y situadas en la misma fila (en el archivo combinado prevalece el nombre de la variable perteneciente al archivo de trabajo).

También es posible incluir en el archivo combinado cualquier variable individual de la lista **Variables desemparejadas** (aunque ya sabemos que estas variables se encuentran en sólo uno de los dos archivos). Para ello:

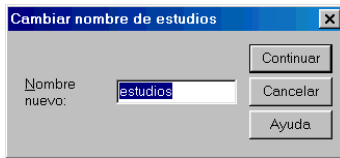
- ▣ Marcar la variable que se desea incluir y pulsar el botón flecha.

Lógicamente, puesto que la variable *desemparejada* sólo se encuentra en uno de los dos archivos, los casos del archivo que no contiene esa variable serán, en el archivo combinado, casos con valor perdido en esa variable.

Cambiar nombre. Previamente a la acción de emparejar dos variables, puede resultar conveniente cambiar el nombre a cualquiera de ellas para que ambas posean el mismo nombre. Para cambiar el nombre de una variable:

- ▶ En la lista **Variables desemparejadas**, seleccionar la variable cuyo nombre se desea cambiar y pulsar el botón **Cambiar nombre...** para acceder al cuadro de diálogo *Cambiar nombre de...* que muestra la figura 6.5.

Figura 6.5. Cuadro de diálogo *Cambiar nombre de...*



Este subcuadro de diálogo permite renombrar la variable seleccionada (*estudios* en nuestro ejemplo). Por supuesto, una variable puede ser renombrada independientemente de que deseemos o no emparejarla con otra.

Variables en el nuevo archivo de datos de trabajo. Contiene un listado de las variables que pasarán a formar parte del nuevo archivo combinado. Por defecto, este listado incluye las variables que tienen el mismo nombre y formato en ambos archivos (de trabajo y externo). Las variables de este listado que no se desee incluir en el nuevo archivo combinado pueden eliminarse seleccionándolas y desplazándolas a la lista **Variables desemparejadas**.

- **Indicar origen del caso como variable.** Esta opción permite crear una *variable indicador* para identificar a qué archivo (de trabajo o externo) pertenecía originalmente cada caso del nuevo archivo combinado. Esta variable *indicador* toma el valor 0 para los casos del archivo de trabajo y el valor 1 para los casos del archivo externo. El nombre por defecto para esta variable *indicador* es *origen01*, pero es posible introducir cualquier nombre válido en el cuadro de texto destinado a tal efecto.

Añadir variables

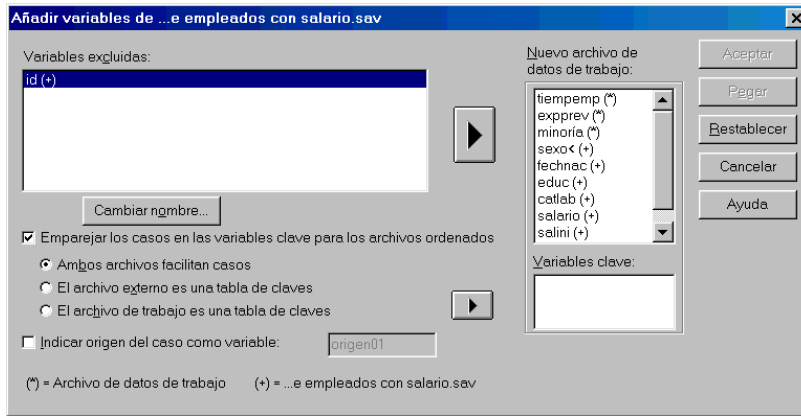
Nota: Para ilustrar el procedimiento **Fundir archivos** > **Añadir variables** hemos partido en dos el archivo *Datos de empleados*. Las variables *id*, *sexo*, *fechnac*, *educ*, *catlab*, *salario* y *salini* las hemos archivado en *Datos de empleados con salario*; y las variables *id*, *tiempemp*, *expprev* y *minoría* las hemos archivado en *Datos de empleados sin salario*. Así pues, aunque ambos archivos contienen los mismos casos, contienen distintas variables (excepto por lo que se refiere a la variable *id*, que se encuentra en ambos archivos).

Para combinar dos archivos con los mismos casos pero con distintas variables es necesario que ambos archivos sean archivos de datos con formato SPSS y que los casos estén en el mismo orden. Además, si se utiliza una o más variables como criterio de emparejamiento, los casos deben estar ordenados de forma ascendente en esa o esas variables (en ambos archivos).

Para fundir dos archivos añadiendo variables debemos comenzar abriendo en el *Editor de datos* uno de los archivos que deseamos combinar (a este primer archivo lo llamaremos **archivo de trabajo**). Abrir el archivo *Datos de empleados con salario* y, tras esto:

- ▶ Seleccionar la opción **Fundir archivos** > **Añadir variables...** del menú **Datos** para acceder al cuadro de diálogo *Añadir variables: Leer archivo* (idéntico al que muestra la figura 6.3).

Este cuadro de diálogo, que contiene un listado de los archivos de datos de la carpeta por defecto, permite seleccionar el **archivo externo** que deseamos combinar con el de trabajo. Al archivo resultante de la fusión lo llamaremos **archivo combinado**. Seleccionando el archivo *Datos de empleados sin salario* y pulsando el botón **Abrir** accedemos al cuadro de diálogo *Añadir variables de...* que muestra la figura 6.6.

Figura 6.6. Cuadro de diálogo *Añadir variables de...*

En el ejemplo de la figura 6.6, el archivo de datos externo seleccionado es *Datos de empleados sin salario*, lo cual queda reflejado en el título del cuadro de diálogo y en la parte inferior izquierda, precedido del signo “+”.

Variables excluidas. Esta lista muestra las variables que no formarán parte del archivo combinado. Incluye, por defecto, las variables del archivo externo cuyo nombre coincide con el de alguna variable del archivo de trabajo. Si se quiere incluir en el archivo combinado el contenido de una variable con nombre duplicado, es necesario renombrar esa variable (botón **Cambiar nombre...**) y añadirla a la lista **Nuevo archivo de datos de trabajo** mediante el botón flecha. Al igual que antes, las variables del archivo de trabajo aparecen acompañadas de un asterisco (*), mientras que las variables del archivo externo van acompañadas de un signo más (+).

Nuevo archivo de datos de trabajo. Esta lista muestra las variables que pasarán a formar parte del archivo combinado. Por defecto, el listado recoge: 1) todas las variables del archivo de trabajo y 2) las variables del archivo externo cuyo nombre no está duplicado en el de trabajo.

☐ **Emparejar los casos en las variables clave para los archivos ordenados.** Si uno de los dos archivos posee más casos que el otro, es posible utilizar una o más variables clave para

emparejar correctamente los casos de ambos archivos. Esta variable clave debe tener el mismo nombre en los dos archivos y los casos deben estar, en ambos archivos, ordenados de forma ascendente según esa variable clave (si se utiliza más de una variable clave, los casos deben estar ordenados de forma ascendente en el conjunto de variables claves utilizadas).

Puesto que las variables clave deben ser variables duplicadas, aparecerán en la lista **Variables excluidas**. Para utilizar una variable clave, es necesario marcar la opción **Emparejar los casos en las variables clave para los archivos ordenados**, seleccionar las variables que se van a utilizar como claves, y trasladarlas a la lista **Variables clave** mediante el correspondiente botón flecha. Hecho esto, podemos elegir entre tres métodos diferentes de emparejamiento de casos:

- ☐ **Ambos archivos facilitan casos.** Esta opción supone que existe una correspondencia uno a uno entre los casos de ambos archivos. Es decir, se supone que cada caso posee un valor único en la(s) variable(s) clave(s). Si existen dos o más casos con el mismo valor en la(s) variable(s) clave(s), el SPSS los empareja en orden secuencial y ofrece un mensaje advirtiéndolo de tal circunstancia.
 - ☐ **El archivo externo es una tabla de claves.** El archivo externo actúa como una tabla de claves. Cada caso del archivo externo puede emparejarse con más de un caso del archivo de trabajo.
 - ☐ **El archivo de trabajo es una tabla de claves.** El archivo de trabajo actúa como una tabla de claves. Cada caso del archivo de trabajo puede emparejarse con más de un caso del archivo de trabajo.
- ☐ **Indicar origen del caso como variable.** Esta casilla permite crear una *variable indicador* para identificar a qué archivo (de trabajo o externo) pertenecía originalmente cada caso del nuevo archivo combinado. Esta variable *indicador* toma el valor 0 para los casos del archivo de trabajo ausentes del archivo externo y el valor 1 para los casos del archivo externo ausentes del archivo de trabajo. El nombre por defecto para esta variable *indicador* es *origen01*, pero es posible asignar cualquier nombre válido escribiéndolo en el cuadro de texto destinado a tal efecto.

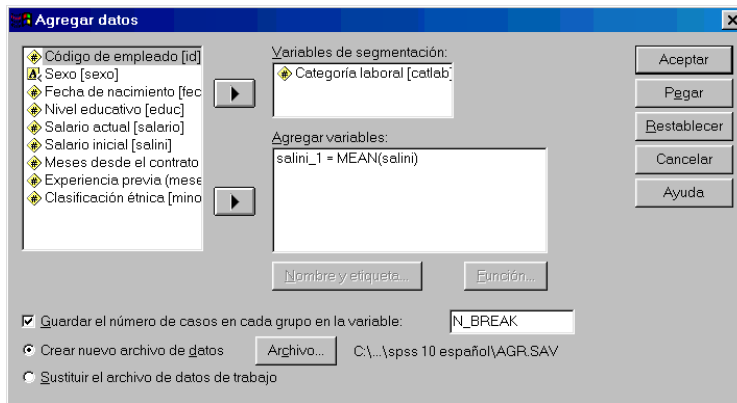
Agregar datos

Agregar datos consiste en agrupar varios casos en uno solo. La opción **Agregar** crea un nuevo archivo de datos en el que cada nuevo caso representa a un conjunto de casos del archivo de datos original. Un archivo agregado tiene, por tanto, menos casos que el archivo original.

Más adelante, en este mismo capítulo, estudiaremos la opción *Ponderar casos*, la cual, según veremos permite hacer justamente lo contrario de lo que hace la opción *Agregar casos*. Para agregar casos:

- ▣ Seleccionar la opción **Agregar...** del menú **Datos** para acceder al cuadro de diálogo *Agregar datos* que muestra la figura 6.7.

Figura 6.7. Cuadro de diálogo *Agregar datos*.



Variables de segmentación. Los casos del archivo original son agrupados a partir de los niveles de una o más variables de segmentación. En el ejemplo de la figura 6.7 hemos utilizado la variable *catlab* (categoría laboral), lo que significa que todos los sujetos que tengan la misma categoría laboral (el mismo valor en *catlab*) pasarán a formar un único caso en el nuevo archivo de datos agregados. Las variables de segmentación pueden ser numéricas o de cadena.

Agregar variables. Las variables del nuevo archivo se obtienen a partir de variables del archivo original. El valor de cada nuevo caso agregado en estas nuevas variables es el resultado de

aplicar alguna **función** (ver más abajo) a las variables del archivo original. En el ejemplo de la figura 6.7 hemos decidido que el nuevo archivo contenga una variable: *salini_1* (el nombre es automáticamente asignado por el SPSS; ver, más abajo, **Nombre y etiquetas...**). La variable *salini_1* será el resultado de obtener, para cada nuevo caso agregado, la media aritmética (MEAN) de la variable *salini* en todos los casos de cada segmento definido por *catlab*.

- ☐ **Guardar el número de casos de cada grupo en la variable....** Al marcar esta opción, el nuevo archivo crea una variable que contiene el número de casos de cada segmento. Esta variable recibe, por defecto, el nombre *N_BREAK*. Si se desea asignarle otro nombre, basta con introducirlo en correspondiente cuadro de texto.

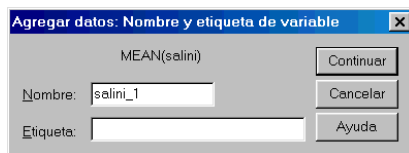
El SPSS ofrece dos alternativas en lo referente a la ubicación del nuevo archivo de datos agregados:

- ☐ **Crear nuevo archivo de datos.** Crea un nuevo archivo con los datos agregados y le asigna el nombre *AGGR.SAV*. Es la opción por defecto. Para crear un nuevo archivo:
 - ☒ Pulsar el botón **Archivo...** para acceder a un subcuadro de diálogo que permite asignar al nuevo archivo agregado un nombre y una ruta diferentes.
- ☐ **Sustituir el archivo de datos de trabajo.** Alternativamente, en lugar de crear un nuevo archivo y almacenarlo en disco, podemos decidir que el nuevo archivo con los datos agregados pase a ser el archivo de trabajo (sustituyendo al archivo de trabajo actual).

Nombre y etiquetas. Las nuevas variables son nombradas, por defecto, añadiendo “_1” a los seis primeros caracteres de la variable original. En la figura 6.8, la variable *salini* ha sido renombrada como *salini_1*. Si se desea cambiar el nombre asignado por defecto:

- ☒ Pulsar el botón **Nombre y etiquetas...** para acceder al cuadro de diálogo *Agregar datos: Nombre y etiqueta de variable* que muestra la figura 6.8.

Figura 6.8. Subcuadro de diálogo *Agregar datos: Nombre y etiqueta de variable*.



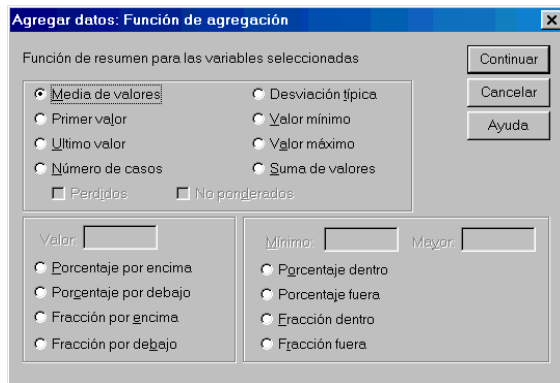
Nombre. Para cambiar el nombre asignado por defecto, escribir el nuevo nombre en este cuadro de texto. El nombre debe comenzar con una letra y no exceder de 8 caracteres (ver capítulo 4, apartado *Definir variables: asignar nombre a una variable*).

Etiqueta. Permite asignar una etiqueta descriptiva de hasta 120 caracteres.

Función. Si no se indica otra cosa, el SPSS asume que la función estadística que se desea utilizar es la media aritmética. No obstante, es posible elegir entre una gran variedad de funciones diferentes. Para seleccionar una función distinta de la media aritmética:

- ▶ Pulsar el botón **Función...** para acceder al subcuadro de diálogo *Agregar datos: Función de agregación* que muestra la figura 6.9.

Figura 6.9. Subcuadro de diálogo *Agregar datos: Función de agregación*.



Este subcuadro de diálogo permite seleccionar varias funciones de forma rápida y sencilla. Todas ellas están apropiadamente descritas en la ayuda contextual que ofrece el SPSS con el botón secundario del ratón. Las funciones se calculan para cada segmento definido por la(s) variable(s) de agrupación, es decir para cada nuevo caso del archivo agregado. Estas funciones aparecen agrupadas en tres bloques:

- Algunos estadísticos descriptivos: la media aritmética, el primer valor, el último valor, el número de casos, la desviación típica, el valor más pequeño, el valor más grande y la suma de valores. Al activar la función *número de casos* es posible decidir si lo que

se desea contar es el número de casos con valores perdidos (marcando la opción *Perdidos*) y si el recuento debe hacerse teniendo en cuenta la ponderación de casos, si existiese (activando y desactivando la opción *No ponderados*).

- El porcentaje o fracción de casos que se encuentran por encima o por debajo de un determinado valor.
- El porcentaje o fracción de casos que se encuentran dentro o fuera de un rango de valores determinado.

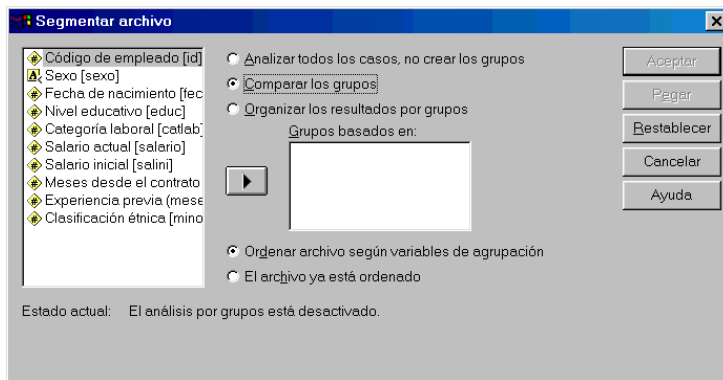
La función elegida en este subcuadro de diálogo aparece en la lista **Agregar variables** del cuadro de diálogo *Agregar datos* (figura 6.7), junto al nombre de la nueva variable.

Segmentar archivo

Segmentar un archivo consiste en dividirlo en subgrupos. Los análisis estadísticos que se llevan a cabo mientras un archivo se encuentra segmentado se repiten para cada grupo resultante de la segmentación. Para segmentar un archivo:

- ▶ Seleccionar la opción **Segmentar archivo...** del menú **Datos** para acceder al cuadro de diálogo *Segmentar archivo* que muestra la figura 6.10.

Figura 6.10. Cuadro de diálogo *Segmentar archivo*.



- **Analizar todos los casos, no crear los grupos.** Esta opción está activa mientras no se solicita segmentar el archivo. Una vez segmentado el archivo, permanece segmentado durante toda la sesión hasta que se vuelve a activar esta opción (o hasta que se realiza una segmentación diferente).
- **Comparar los grupos.** Al marcar esta opción se activa la segmentación y los análisis solicitados a partir de este momento se repiten para cada segmento. Si se solicitan varios análisis, la salida los ofrece organizados de la siguiente manera: el primer análisis para todos los grupos, el segundo análisis para todos los grupos, etc.
- **Organizar los resultados por grupos.** Al marcar esta opción se activa la segmentación y los análisis solicitados a partir de este momento se repiten para cada segmento. Si se

solicitan varios análisis, la salida los ofrece organizados de la siguiente manera: todos los análisis para el primer grupo, todos los análisis para el segundo grupo, etc.

Grupos basados en. Para segmentar el archivo de datos es necesario seleccionar una o más variables de segmentación. Para ello:

- ▣ Seleccionar la(s) variable(s) en la lista de variables del archivo de datos y pulsar el botón flecha.

Las variables de segmentación pueden ser numéricas o de cadena. No es posible seleccionar más de 8 variables (a efectos de este límite, cada 8 caracteres de una variable de cadena larga cuentan como una variable). Al seleccionar más de una variable de segmentación, los segmentos se establecen siguiendo el orden de selección de las variables.

- **Ordenar archivo según variables de segmentación.** Ordena por segmentos los casos del archivo de datos. Es la opción por defecto. Siempre es conveniente ordenar el archivo de datos por las variables utilizadas en la segmentación. En el caso de que se utilice más de una variable de segmentación, no ordenar el archivo puede conducir a segmentos inconsistentes.
- **El archivo ya está ordenado.** Impide que los casos sean reordenados.

Cuando el archivo se encuentra segmentado, la barra de estado de la ventana principal muestra el mensaje *Segmentado*.

Seleccionar casos

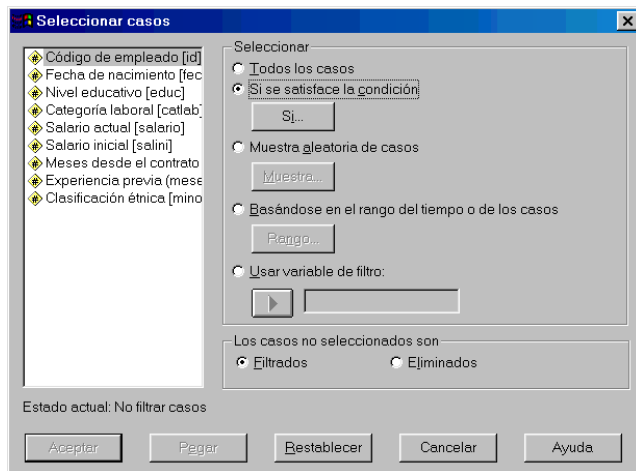
En ocasiones puede interesar centrar el análisis en sólo un grupo de casos que cumplan determinada condición. En otras ocasiones puede que sólo interese analizar una muestra aleatoria del total de casos del archivo de datos.

El SPSS permite seleccionar un conjunto de casos utilizando diferentes criterios: valores o rangos de valores de una variable, números de registro, expresiones aritméticas y lógicas, funciones matemáticas, etc. La selección de casos es una opción a la que todo usuario del SPSS termina encontrando gran utilidad. Para seleccionar casos:

- ▶ Seleccionar la opción **Seleccionar casos...** del menú **Datos** para acceder al cuadro de diálogo *Seleccionar casos* que muestra la figura 6.11.

El recuadro **Seleccionar** contiene todas las opciones de selección de casos disponibles en el SPSS.

Figura 6.11. Cuadro de diálogo *Seleccionar casos*.



Para seleccionar un grupo de casos puede optarse por cualquiera de las siguientes cuatro alternativas:

- **Todos los casos.** Selecciona todos los casos del archivo de datos. Es la opción por defecto. Cuando la selección de casos está activa, esta opción la anula.
- **Si se satisface la condición.** Permite seleccionar casos que satisfagan una determinada condición. Para establecer la condición:

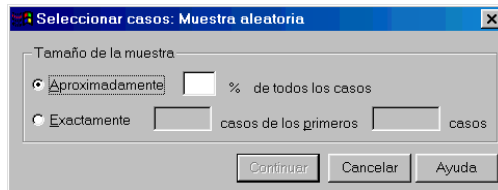
- ▢ Pulsar el botón **Si...** para acceder al cuadro de diálogo *Seleccionar casos si*, similar al ya estudiado en el capítulo 5, figura 5.4.

Este cuadro de diálogo permite utilizar expresiones condicionales basadas en funciones aritméticas, lógicas, relacionales, estadísticas, etc. (Ver, en el capítulo 5, el apartado *Calcular: Expresiones condicionales*).

- **Muestra aleatoria de casos.** Esta opción selecciona aleatoriamente un porcentaje o un número de casos. Para obtener una muestra aleatoria:

- ▢ Pulsar el botón **Muestra...** para acceder al subcuadro de diálogo *Seleccionar casos: Muestra aleatoria* que recoge la figura 6.12.

Figura 6.12. Subcuadro de diálogo *Seleccionar casos: Muestra aleatoria*.



Tamaño de muestra. Existen dos formas de obtener una muestra aleatoria:

- **Aproximadamente ___ % de todos los casos.** Selecciona, aproximadamente, el porcentaje de casos indicado.
- **Exactamente ___ casos de los primeros ___ casos.** Selecciona exactamente el número de casos indicado. Selecciona los casos de entre los n primeros,

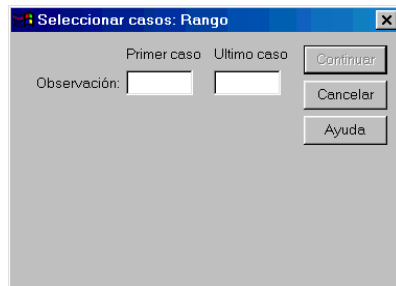
siendo n un número definido por el usuario. Este número debe ser menor o igual que el número de casos del archivo de datos. Si es mayor, el tamaño de la muestra aleatoria será menor que el indicado.

Cada vez que el SPSS genera una muestra aleatoria utiliza una *Semilla de aleatorización* diferente, lo que da lugar a muestras aleatorias diferentes. Si se desea replicar la misma muestra es necesario establecer la misma semilla de aleatorización (ver, en el capítulo 5, el apartado *Semilla de aleatorización*).

- **Basándose en el rango del tiempo o de los casos.** Permite seleccionar un rango de casos a partir del número de registro (fila) que ocupan en el *Editor de datos*. Cuando los casos constituyen una serie temporal y existen variables *fecha*, es posible establecer un rango de fechas u horas para seleccionar casos. Para ello:

- ▶ Pulsar el botón **Rango...** para acceder al subcuadro de diálogo *Seleccionar casos: Rango*: *Rango* que muestra la figura 6.13 y fijar el rango de valores que se desea seleccionar utilizando las cajas **Primer caso** y **Último caso**.

Figura 6.13. Subcuadro de diálogo *Seleccionar casos: Rango*.



- **Usar variable de filtro.** Esta opción permite utilizar cualquier variable numérica del archivo de datos para efectuar la selección de casos o, incluso, para eliminar casos de forma selectiva. Son seleccionados los casos cuyo valor en esa variable es distinto de cero. Los casos con valor cero son excluidos.

Los casos no seleccionados son. Los casos no seleccionados pueden tratarse de dos formas diferentes:

- **Filtrados.** Los casos no seleccionados no son incluidos en los análisis, pero permanecen en el archivo de datos. Pueden volverse a utilizar durante la misma sesión si se desactiva la selección de casos.

Al seleccionar una muestra aleatoria o al seleccionar casos utilizando una expresión condicional, la opción **Filtrados** crea una variable nueva llamada *filter_\$* con “unos” para los casos seleccionados y “ceros” para los no seleccionados. Al efectuar una nueva selección, el contenido de la variable *filter_\$* cambia para reflejar el resultado de la nueva selección. Si deseamos guardar la variable *filter_\$* y utilizarla como variable de filtro en otro momento (por ejemplo, para no tener que invertir tiempo en definir de nuevo una expresión condicional compleja que nos ha costado bastante trabajo construir), podemos cambiarle el nombre y guardar el archivo de datos con la nueva variable.

Los casos no seleccionados pueden identificarse en el archivo de datos por una línea diagonal (a modo de tachadura) sobre el número de caso, es decir, sobre la cabecera de la fila correspondiente al caso no seleccionado.

- **Eliminados.** Los casos no seleccionados son eliminados del archivo de datos. Esto permite ganar tiempo de procesamiento. Los casos eliminados pueden recuperarse abriendo de nuevo el archivo de datos sin haberlo salvado tras la selección (es decir, la eliminación de casos sólo es permanente si, una vez eliminados los casos no seleccionados, se guarda el archivo de datos sin cambiarle el nombre).

Cuando la selección de casos se encuentra activa, la barra de estado de la ventana principal muestra el mensaje *Filtrado*.

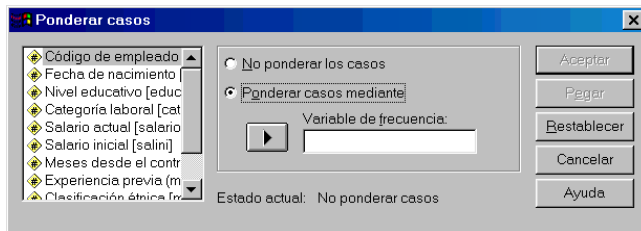
Ponderar casos

Ponderar casos consiste en hacer que un registro (caso) represente a más de un caso. El resultado de la ponderación es justamente el inverso de la agregación.

Ponderar casos exige utilizar una *variable de ponderación* que es justamente la que contiene los pesos que serán asignados a cada caso. Para ponderar casos:

- ▣ Seleccionar la opción **Ponderar casos...** del menú **Datos** para acceder al cuadro de diálogo *Ponderar casos* que muestra la figura 6.14.

Figura 6.14. Cuadro de diálogo *Ponderar casos*.



La lista de variables muestra únicamente las variables con formato numérico. Para seleccionar una variable de ponderación:

- ▣ Activar la opción **Ponderar casos mediante**.
- ▣ Seleccionar, en la lista de variables del archivo de datos, la variable que se desea utilizar como variable de ponderación.
- ▣ Pulsar el botón flecha para trasladar esa variable al cuadro **Variable de frecuencia**.

Una vez activada la ponderación, ésta permanece activa hasta que se cambia de variable de ponderación o hasta que se marca la opción **No ponderar casos**.

La parte inferior del cuadro de diálogo, **Estado actual**, nos informa sobre si la ponderación está activa o no. Además, cuando la ponderación se encuentra activa, la barra de estado de la ventana principal muestra el mensaje *Ponderado*.

La ponderación de casos posee una utilidad especial relacionada con la reproducción de tablas de contingencias. Una tabla de contingencias es una forma concreta de organizar las frecuencias referidas a dos o más variables categóricas. La tabla 1 muestra un ejemplo de tabla

de contingencias con las frecuencias (número de casos) resultantes de cruzar las variables *sexo* y *tabaco*.

Tabla 6.1. Tabla de contingencias de *sexo* por *tabaco*.

		Tabaco			
		Fumadores (1)	No fumadores (2)	Exfumadores (3)	
Sexo	Varones (1)	360	600	240	1200
	Mujeres (2)	340	300	160	800
		700	900	400	2.00

Esta tabla se ha obtenido utilizando dos variables y una muestra de 2.000 casos. Si nos encontramos en un informe con esta tabla y deseamos reproducirla en el *Editor de datos* para efectuar nuestros propios análisis, no necesitamos crear un archivo de datos con 2.000 casos. Gracias a la opción *ponderar*, sólo necesitaremos crear tantos casos como casillas tiene la tabla.

La tabla 6.1 tiene 6 casillas, luego nuestro archivo de datos estará formado por 6 casos. Pero además de crear las dos variables de la tabla (*sexo* y *tabaco*), necesitamos crear una tercera variable (a la que podemos llamar *ncasos*) que recoja las frecuencias de cada casilla. La figura 6.15 muestra cómo hacer esto.

Figura 6.15. *Editor de datos*: reproducción de los datos de la tabla 6.1.



Para reproducir la tabla 6.1 en el *Visor de resultados* a partir de los datos que muestra el *Editor de datos* de la figura 6.15:

- ▶ Seleccionar **Ponderar** en el menú **Datos** para acceder al cuadro de diálogo *Ponderar casos* (figura 6.14).
- ▶ Marcar la opción **Ponderar casos mediante** y trasladar la variable *ncasos* al cuadro **Variable de frecuencia**.
- ▶ Seleccionar **Estadísticos descriptivos > Tablas de contingencia** en el menú **Analizar** y trasladar la variable *sexo* al recuadro **Fila** y la variable *tabaco* al recuadro **Columna**.

Aceptando estas elecciones, el *Visor de resultados* ofrece la tabla de contingencias que muestra la tabla 6.2, la cual, como puede comprobarse, es idéntica a la tabla 6.1. A pesar de que el archivo de datos sólo tiene 6 casos, al ponderar el archivo con la variable *ncasos*, los 6 casos se han convertido en 2.000.

Tabla 6.2. Tabla de contingencias de *sexo* por *tabaco*.

Recuento		Tabaco			Total
		Fumadores	No fumadores	Exfumadores	
Sexo	Varones	360	600	240	1200
	Mujeres	340	300	160	800
Total		700	900	400	2000