

# Capítulo 5

## Transformar datos

En ocasiones, los datos de un archivo serán de tal índole que será posible aplicar directamente sobre ellos el análisis estadístico deseado. Pero esto sólo ocurrirá en una situación ideal; y las situaciones ideales raramente se presentan. Más bien al contrario, lo habitual será encontrarse con archivos de datos que necesitarán ser cuidadosamente preparados antes de poder aplicar con las mínimas garantías cualquier análisis estadístico.

La *preparación* del archivo de datos incluye desde la simple detección y corrección de los posibles errores cometidos al introducir datos, hasta sofisticadas transformaciones (necesarias a veces para llegar a obtener las variables que realmente interesan), pasando por la recodificación de los códigos utilizados para los valores de una o más variables, o la creación de nuevas variables a partir de otras ya existentes.

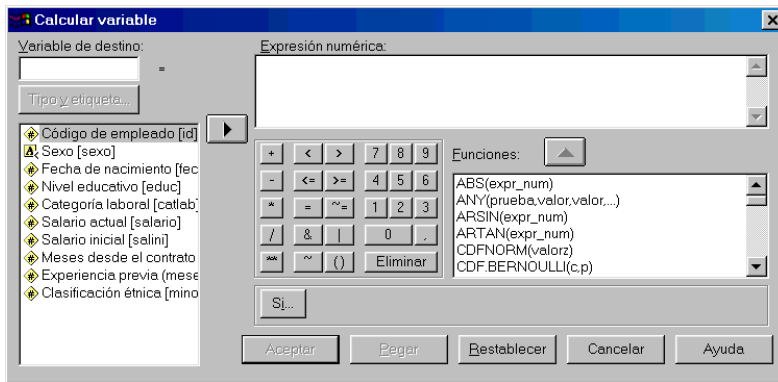
El menú **Transformar** de la barra de menús principal incluye una serie de opciones que permiten efectuar diferentes tipos de transformaciones, desde las más simples a las más complejas. Trataremos en este capítulo todas esas opciones.

## Calcular

El SPSS incluye varias opciones para crear variables nuevas a partir de otra u otras existentes poniendo a disposición del usuario una gran cantidad de funciones diferentes. Para crear una variable nueva:

- ▣ Seleccionar la opción **Calcular...** del menú **Transformar** para acceder al cuadro de diálogo *Calcular variable* (figura 5.1).

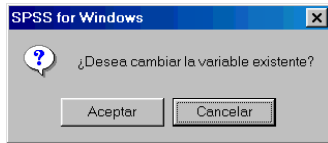
**Figura 5.1.** Cuadro de diálogo *Calcular variable*.



## Variable de destino

El cuadro de texto **Variable de destino** permite introducir el nombre de la variable que recibirá los valores calculados. El nombre de esta variable puede ser nuevo o puede ser el de una variable ya existente. Si se propone un nombre nuevo, éste debe respetar las reglas de los nombres de variable (ver, en el capítulo 4, el apartado *Definir variables*). Si el nombre propuesto coincide con el de una variable ya existente, al pulsar el botón **Aceptar** aparece un mensaje de aviso (figura 5.2) solicitando confirmar o cancelar la acción.

**Figura 5.2.** Cuadro de advertencia de variable duplicada.

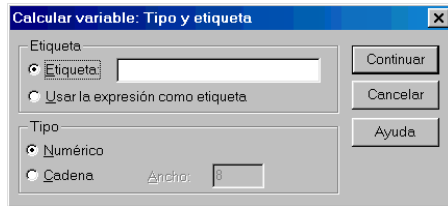


## Tipo de variable y etiqueta

A la variable de destino se le asigna, por defecto, formato numérico (siempre que la plantilla *PREDETERMINADA*, que es la que actúa por defecto, no haya sido modificada). No obstante, el cuadro de diálogo **Calcular** permite controlar el tipo de formato de las nuevas variables. Para cambiar el formato de la nueva variable y/o asignarle una etiqueta:

- ▶ Pulsar el botón **Tipo y etiqueta...** para acceder al subcuadro de diálogo *Calcular variable: Tipo y etiqueta* que muestra la figura 5.3.

**Figura 5.3.** Subcuadro de diálogo *Calcular variable: Tipo y etiqueta*.



El recuadro **Etiqueta** permite asignar una etiqueta descriptiva a la nueva variable:

- **Etiqueta.** Admite una etiqueta descriptiva de hasta 120 caracteres.
- **Usar la expresión como etiqueta.** Al seleccionar esta opción, el SPSS asigna como etiqueta los primeros 110 caracteres de la expresión numérica.

El recuadro **Tipo** sirve para cambiar el formato de la nueva variable:

- **Numérico.** Formato por defecto.
- **Cadena.** Asigna formato de cadena. Si se opta por este formato, es necesario especificar el ancho de la cadena en la caja **Ancho**.

## Expresión numérica

En este cuadro debe escribirse la expresión numérica encargada de generar los valores de la variable de destino. Una expresión numérica es una expresión matemática similar a la que puede construirse con una calculadora de bolsillo, con la diferencia de que, además de constantes y operadores aritméticos, admite nombres de variables ya existentes, operadores relacionales y lógicos, y una gran variedad de funciones matemáticas.

## Calculadora

Para facilitar la tarea de construir la expresión numérica, el cuadro de diálogo *Calcular variable* incluye un teclado de calculadora con números y operadores aritméticos, relacionales y lógicos. Este teclado funciona exactamente igual que el de una calculadora convencional (pero pinchando con el puntero del ratón).

Operadores aritméticos	Operadores relacionales	Operadores lógicos
+ Suma	< Menor que	& “y”
– Resta	> Mayor que	“o”
* Multiplicación	<= Menor o igual que	~ “no”
/ División	>= Mayor o igual que	
** Potencia	= Igual	
( ) Orden de las operaciones	~≠ Distinto	


En relación con los operadores **aritméticos** hay que tener presente, sobre todo, el *orden* en el que operan: en primer lugar se evalúan las funciones, después las potencias, después la multiplicación y la división y, por último, la suma y la resta. Pueden utilizarse los paréntesis para alterar el orden de las operaciones. Los operadores **relacionales** y **lógicos** son útiles, sobre todo, para efectuar transformaciones condicionales (enseguida nos referimos a ellas). Algunos ejemplos:

1.  $2 * (var1 - var2) ** 2$
2.  $(var1 + var2 + var3) / 3$
3.  $var1 * 5 / 100$

La primera expresión calcula el doble de la diferencia al cuadrado entre las variables *var1* y *var2*: cada caso del archivo de datos pasa a tener, en la nueva variable, el doble del valor resultante de elevar al cuadrado la resta de sus puntuaciones en *var1* y *var2*. La segunda expresión calcula la media aritmética de las variables *var1*, *var2* y *var3*: cada caso del archivo de datos pasa a tener, en la nueva variable, la media aritmética de sus puntuaciones en esas tres variables. La tercera expresión calcula el 5 por ciento de la variable *var1*: el valor de cada caso en la nueva variable será el 5 por ciento de su puntuación en *var1*.

## Funciones

La lista del recuadro **Funciones** incluye más de 70 funciones aritméticas, estadísticas, lógicas, etc. La ayuda específica del cuadro de diálogo *Calcular variable* contiene información puntual sobre estas funciones. Además, situando el puntero del ratón sobre una función, al pulsar el botón secundario del ratón aparece un cuadro de ayuda contextual con una descripción de esa función. Para trasladar una función al cuadro de texto **Expresión numérica**:

- ▶ Seleccionar la función y pulsar el botón *flecha*  para trasladarla al cuadro **Expresión numérica**.

Una expresión no está completa hasta que se incluye entre paréntesis el *argumento* de la función seleccionada. Algunos ejemplos:

1. MEAN (var1, var2, var3).
2. ARSIN (ABS(SQRT(Var1))).
3. RV.NORMAL (0, 1).

La primera expresión calcula la media aritmética de *var1*, *var2* y *var3*. La segunda calcula la función arcoseno del valor absoluto de la raíz cuadrada de la variable *var1* (transformación ésta muy utilizada, por ejemplo, cuando las puntuaciones de la variable dependiente de un análisis de varianza son proporciones). La tercera expresión genera una variable aleatoria (*RV = random variable*) distribuida normalmente con media 0 y desviación típica 1.

## Expresiones condicionales

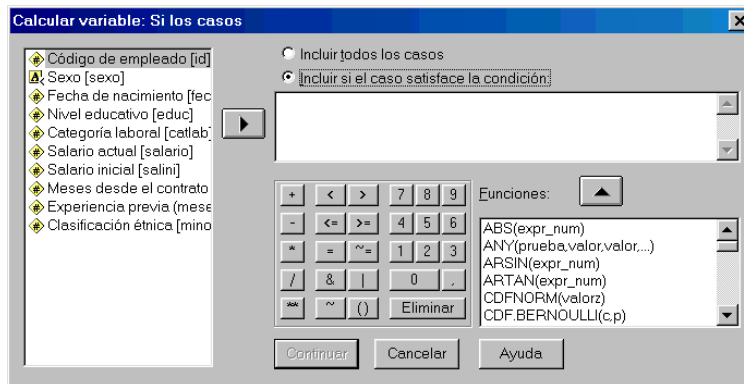
Definida la expresión numérica, el botón **Aceptar** crea una nueva variable en la que *todos los casos válidos* del archivo de datos adoptan el valor resultante de la expresión numérica propuesta. Ahora bien, la expresión numérica no tiene por qué afectar a todos los casos del archivo. De hecho, podemos establecer una condición cualquiera y hacer que las transformaciones propuestas afecten sólo a los casos que cumplan esa condición. Para ello:

- ▶ Pulsar el botón **Si...** del cuadro de diálogo *Calcular variable* (figura 5.1) para acceder al subcuadro de diálogo *Calcular variable: Si los casos* que muestra la figura 5.4.

Este subcuadro de diálogo permite establecer una gran variedad de condiciones para seleccionar sólo los casos que interesen:

- **Incluir todos los casos.** Con esta opción activa, las transformaciones propuestas afectan a todos los casos. Es la opción por defecto.
- **Incluir si el caso satisface la condición.** Esta opción hace que las transformaciones propuestas afecten únicamente a los casos que cumplan la condición establecida. La condición puede incluir nombres de variables, constantes, operadores aritméticos, relacionales y lógicos, y funciones matemáticas. Para construir la condición disponemos de un teclado de calculadora y de una lista de funciones idénticas a los del cuadro de diálogo **Calcular variables**.

**Figura 5.4.** Cuadro de diálogo *Calcular variable: Si los casos*.





Si la variable de destino es una variable nueva, los casos seleccionados (aquellos que cumplen la condición establecida) adoptan en ella el valor resultante de la expresión numérica; y los casos no seleccionados (los que no cumplen la condición establecida) se consideran valores perdidos (aparecen puntos en las casillas del *Editor de datos*).

Si la variable de destino es una variable ya existente, los casos seleccionados adoptan el valor resultante de la expresión numérica y los casos no seleccionados quedan como estaban.

La expresión numérica puede construirse utilizando la lista de variables del archivo de datos, los botones de la calculadora y la lista de funciones. Pero también puede construirse utilizando el teclado directamente, en cuyo caso hay que tener en cuenta unas pocas reglas para no cometer errores:

- Los valores de las variables de cadena deben escribirse entre apóstrofos o entre comillas. Hay que tener especial cuidado en no dejar espacios en blanco donde no deba haberlos.
- Los argumentos de una función deben ir entre paréntesis y, cuando haya más de uno, separados por comas. Puede insertarse un espacio en blanco entre un paréntesis y un argumento, o entre un argumento y otro, pero no es necesario.
- Dentro de una expresión compleja, todas las expresiones simples deben estar completas. Por ejemplo, `edad > 18 & < 30` es una expresión incorrecta, por incompleta. La expresión correcta es: `edad > 18 & edad < 30`.
- El punto es el único separador decimal válido, independientemente de las especificaciones internacionales seleccionadas en Windows.

## Ejemplo (Transformar > Calcular > Si...)

Un sencillo ejemplo nos servirá para formarnos una idea acerca de cómo se crean variables utilizando expresiones condicionales. Vamos a crear la variable *salario2* aplicando a la variable *salario* una subida del 10 % para las *mujeres administrativas* y del 5 % para el resto de los casos. Para ello:

- ▶ Seleccionar la opción **Calcular...** del menú **Transformar** para acceder al cuadro de diálogo *Calcular variable* (figura 5.1).
- ▶ Introducir el nombre *salario2* en el cuadro de texto **Variable de destino**.
- ▶ Introducir en el cuadro de texto **Expresión numérica** una expresión que permita incrementar la variable *salario* en un 10 por ciento; por ejemplo: *salario* \* 1.10.
- ▶ Pulsar el botón **Si...** para acceder al subcuadro de diálogo *Calcular variable: Si los casos* (figura 5.4) y establecer la condición necesaria para aplicar la expresión numérica sólo a las mujeres administrativas: *sexo* = 'm' and *catlab* = 1 (nótese que los valores de las variables de cadena —*sexo* en este ejemplo—van entre comillas o apóstrofes).
- ▶ Pulsar el botón **Continuar** (figura 5.4) y el botón **Aceptar** (figura 5.1) para hacer efectiva esta primera transformación.
- ▶ Volver a entrar en el cuadro de diálogo *Calcular variable* (figura 5.1) e introducir en el cuadro de texto **Expresión numérica** una expresión que permita incrementar la variable *salario* en un 5 por ciento; por ejemplo: *salario* \* 1.05.
- ▶ Pulsar el botón **Si...** para acceder al subcuadro de diálogo *Calcular variable: Si los casos* (figura 5.4) y establecer la condición necesaria para aplicar esta segunda expresión numérica a todos los casos excepto a las mujeres administrativas: *sexo* = 'h' and *catlab* = 1 or *catlab* > 1.
- ▶ Pulsar el botón **Continuar** (figura 5.4) y el botón **Aceptar** (figura 5.1) para hacer efectiva esta segunda transformación.

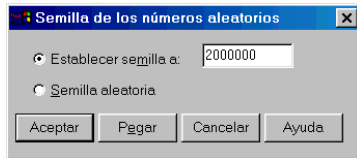
## Semilla de aleatorización

Las funciones del SPSS que utilizan números aleatorios (como, por ejemplo, la función de probabilidad RV.NORMAL o la selección aleatoria de una muestra del archivo de datos) se basan en el generador de números aleatorios del SPSS, el cual siempre comienza a generar una serie aleatoria a partir de un valor inicial llamado **semilla**.

Esta semilla, por defecto, cambia de forma aleatoria cada vez que el SPSS genera una serie aleatoria. Esto significa que las distintas series aleatorias solicitadas al SPSS no serán siempre las mismas (justamente por ser aleatorias). No obstante, existe la posibilidad de replicar una serie aleatoria si se fuerza al generador de números aleatorios a comenzar con la misma semilla. Para ello:

- ▶ Seleccionar la opción **Semilla de aleatorización...** del menú **Transformar** para acceder al cuadro de diálogo *Semilla de los números aleatorios* que muestra la figura 5.5.

**Figura 5.5.** Cuadro de diálogo *Semilla de los números aleatorios*.



- **Establecer semilla a.** Esta opción permite seleccionar como semilla un entero positivo entre 1 y 2.000.000.000. Siempre que se utiliza la misma semilla, se obtiene la misma serie de números aleatorios.
- **Semilla aleatoria.** Marcando esta opción, cada serie aleatoria solicitada es generada a partir de una semilla obtenida de forma aleatoria. De esta forma, las series aleatorias generadas serán, presumiblemente, distintas.

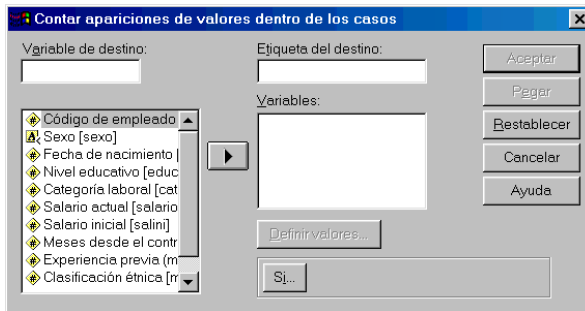
## Contar apariciones

Esta opción sirve para crear variables nuevas a partir del número de veces que se repite uno o más valores determinados en un conjunto de variables. Una variable creada con esta opción contiene, para cada caso, el valor resultante de contar el número de veces que el(los) valor(es) seleccionado(s) aparece(n) en el conjunto de variables seleccionadas.

**Contar apariciones** es útil, por ejemplo, para contar el número de valores perdidos que acumula cada caso del archivo, o para corregir el número de aciertos de los sujetos en preguntas de opción múltiple, etc. Para contar valores:

- ▣ Seleccionar la opción **Contar apariciones...** del menú **Transformar** para acceder al cuadro de diálogo *Contar apariciones de valores dentro de casos* que muestra la figura 5.6.

**Figura 5.6.** Cuadro de diálogo *Contar apariciones de valores dentro de los casos*.



**Variable de destino.** Se refiere al nombre de la nueva variable: la variable que recogerá el resultado del recuento. El nombre de la variable de destino puede ser nuevo o puede ser el de una variable ya existente. Si el nombre es nuevo, éste debe respetar las reglas de los nombres de variable (ver, en el capítulo 4, el apartado *Definir variables*). Si el nombre de la variable de destino es el de una variable ya existente, al pulsar el botón **Aceptar** aparece un mensaje de aviso (ver figura 5.2) solicitado confirmar o cancelar la acción. Lógicamente, la variable de destino debe tener formato numérico.

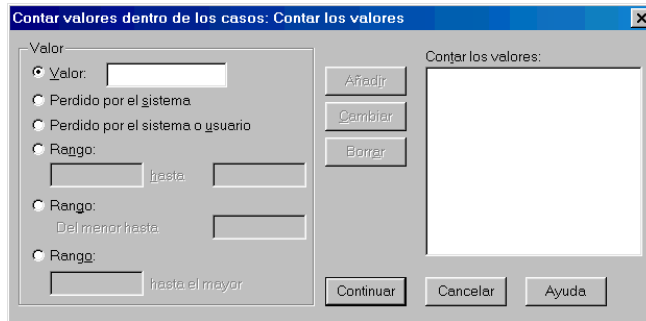
**Etiqueta del destino.** Permite asignar una etiqueta descriptiva de hasta 120 caracteres a la variable de destino. Si la variable de destino ya existe y tiene etiqueta, la etiqueta aparece en este cuadro de texto.

**Variables.** Las variables seleccionadas, es decir, las variables sobre las que deseamos efectuar el recuento, deben trasladarse a esta lista. Estas variables pueden ser numéricas o de cadena, pero no ambas cosas.

**Definir valores.** Para indicar el valor o valores que se desea contar:

- ▶ Pulsar el botón **Definir valores...** para acceder al subcuadro de diálogo *Contar valores dentro de los casos: Contar los valores* que muestra la figura 5.7.

**Figura 5.7.** Subcuadro de diálogo *Contar valores dentro de los casos: Contar los valores*.



El recuadro **Valor** contiene varias opciones para especificar el valor o valores que se desea contar. Permite seleccionar valores individuales, rangos de valores o una combinación de ambas cosas:

- **Valor.** Para contar el número de veces que aparece un valor concreto.
- **Perdido por el sistema.** Para contar el número de valores perdidos definidos por el sistema. En la lista de valores aparece SYSMIS.
- **Perdido por el sistema o usuario.** Para contar el número de valores perdidos de cualquier tipo (definidos por el sistema o definidos por el usuario). En la lista de valores aparece MISSING.

- **Rango (...hasta...)**. Para contar el número de veces que aparecen valores comprendidos entre los límites del rango definido. No disponible para variables de cadena.
- **Rango (Del menor hasta...)**. Para contar el número de veces que aparecen valores comprendidos entre el valor más pequeño y el valor especificado. No disponible para variables de cadena.
- **Rango (...hasta el mayor)**. Para contar el número de veces que aparecen valores comprendidos entre el valor especificado y el mayor. No disponible para variables de cadena.

La función **Contar** efectúa, para cada caso del archivo, un recuento del número de veces que se repite, en el conjunto de variables seleccionadas, cualquier valor de los añadidos a la lista. Para construir la lista de valores que se desea contar:

- ▶ Seleccionar la opción deseada en el recuadro **Valor** y, en caso necesario, introducir el valor o valores en los correspondientes cuadros de texto.
- ▶ Desplazar a la lista **Contar los valores**, mediante el botón **Añadir**, el valor o valores definidos.
- ▶ Utilizar los botones **Cambiar** y **Borrar** para modificar o eliminar, respectivamente, valores previamente añadidos.

La función **Contar**, al igual que la función calcular, puede afectar a todos los casos del archivo o a sólo un conjunto de casos que cumplan determinada condición. Si se desea establecer alguna condición:

- ▶ El botón **Si...** (ver figura 5.6) conduce el subcuadro de diálogo *Contar apariciones: Si los casos* (idéntico al de la figura 5.4), el cual permite establecer las condiciones que deben cumplir los casos seleccionados.

## Recodificar

La opción **Recodificar** del menú **Transformar** permite cambiar los códigos asignados a los valores de una variable. La recodificación resulta especialmente útil para agrupar en un único valor diferentes valores de una variable, y para transformar variables con formato de cadena en variables con formato numérico.

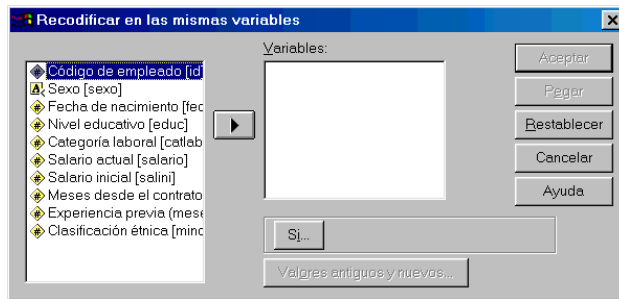
Puede efectuarse sobre *las mismas variables* (cambiando los códigos de una variable existente sin cambiar su nombre) o sobre *variables distintas* (creando una variable nueva a partir de otra ya existente).

## Recodificar en las mismas variables

Para cambiar los códigos de una variable sin cambiar su nombre:

- ▶ Seleccionar la opción **Recodificar > En las mismas variables...** del menú **Transformar** para acceder al cuadro de diálogo *Recodificar en las mismas variables* (figura 5.8).

**Figura 5.8.** Cuadro de diálogo *Recodificar en las mismas variables*.



**Variables.** Las variables cuyos códigos se desea cambiar deben ser trasladadas a esta lista. Para trasladar variables:

- ▶ Seleccionar la variable en la lista de variables del archivo de datos y desplazarla a la lista **Variables** mediante el botón flecha. Puede llevarse más de una variable si es que interesa efectuar *la misma recodificación* a más de una variable.

En la recodificación pueden utilizarse tanto variables numéricas como variables de cadena, pero no al mismo tiempo. Es decir, todas las variables llevadas a la lista **Variables** deben tener el mismo tipo de formato. Al trasladar la primera variable, si es numérica, la lista **Variables** cambia su nombre a **Variables numéricas**; si esa primera variable es de cadena, el nombre de la lista cambia a **Variables de cadena**.

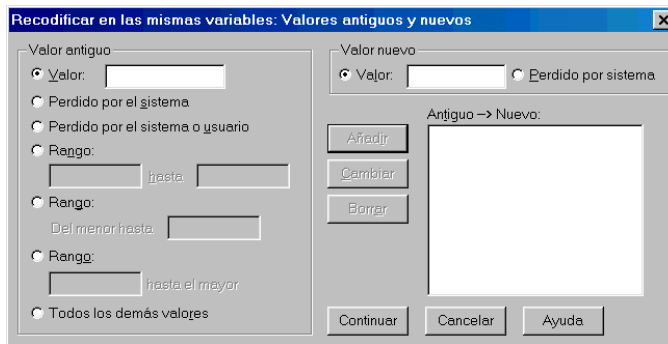
**Si...** La recodificación puede efectuarse de forma condicional, es decir, de forma que sólo afecte a los casos que cumplan determinada condición. Para establecer una condición:

- ▶ Pulsar el botón **Si...** para acceder al subcuadro de diálogo *Recodificar en las mismas variables: Si los casos* (idéntico al de la figura 5.4). Este subcuadro contiene todas las opciones necesarias para efectuar transformaciones condicionales.



**Valores antiguos y nuevos...** Una vez seleccionadas las variables que se van a recodificar, es necesario indicar la recodificación concreta que se desea llevar a cabo. Para ello, el botón **Valores antiguos y nuevos...** conduce al subcuadro de diálogo *Recodificar en las mismas variables: Valores antiguos y nuevos* que muestra la figura 5.9.

**Figura 5.9.** Subcuadro de diálogo *Recodificar en las mismas variables: Valores antiguos y nuevos*.



**Valor antiguo.** Las opciones de este recuadro permiten especificar el valor o valores de la variable original que se desea recodificar. Ofrece varias alternativas para facilitar la identificación del valor o valores antiguos:

- **Valor.** Un valor individual.
- **Perdido por el sistema.** Valores perdidos definidos por el sistema (aparecen como SYSMIS en la lista de valores). Esta opción no está disponible con variables de cadena.
- **Perdido por el sistema o usuario.** Valores perdidos de cualquier tipo: definidos por el sistema o definidos por el usuario (aparecen como MISSING en la lista de valores).
- **Rango (... hasta ...).** Rango de valores comprendidos entre los dos valores indicados. Esta opción no está disponible con variables de cadena.
- **Rango (Del menor hasta ...).** Rango comprendido entre el valor más pequeño de la variable y el valor indicado. Esta opción no está disponible con variables de cadena.

- **Rango (... hasta el mayor).** Rango comprendido entre el valor indicado y el valor más grande de la variable. Esta opción no está disponible con variables de cadena.
- **Todos los demás valores.** Todos los valores de la variable original no definidos previamente (aparece como ELSE en la lista de valores).

**Valor nuevo.** En este recuadro debe especificarse el nuevo valor que se desea asignar al valor o valores antiguos. Ofrece las siguientes alternativas:

- **Valor.** El valor introducido en este cuadro de texto sustituye al valor o valores antiguos recién definidos. (Si se trata de una variable de cadena, el valor introducido en este cuadro de texto aparece entre apóstrofes o comillas al trasladarlo a la lista de valores: por tanto, no hay que poner apóstrofes o comillas a los valores de las variables de cadena, como ocurre, por ejemplo, en los cuadros de diálogo que permiten definir una expresión numérica).
- **Perdido por el sistema.** Asigna un valor perdido definido por el sistema (aparece como SYSMIS en la lista de valores).

Para cada valor o rango de valores que se desee recodificar:

- ▶ Indicar el valor o rango de valores antiguos que se desea recodificar utilizando alguna de las opciones del recuadro **Valor antiguo**.
- ▶ Especificar el nuevo código en el cuadro de texto **Valor nuevo**.
- ▶ Pulsar el botón **Añadir** para trasladar la recodificación recién definida a la lista **Antiguo → Nuevo**.
- ▶ Utilizar los botones **Cambiar** y **Borrar** para modificar o eliminar, respectivamente, recodificaciones previamente añadidas.

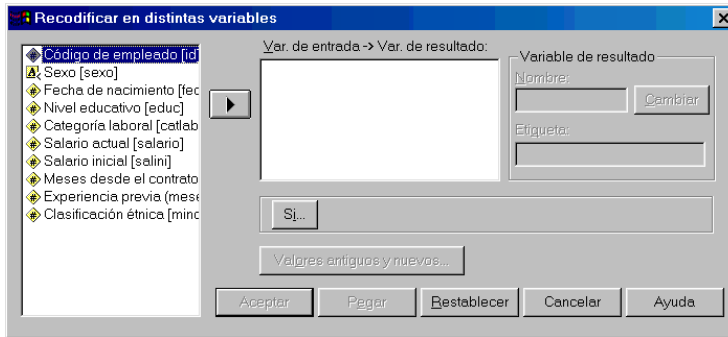
Debe tenerse presente que, aunque es posible asignar *el mismo valor nuevo a más de un valor antiguo* (lo cual puede resultar útil, por ejemplo, para fundir en un sólo valor nuevo valores antiguos no consecutivos), no es posible asignar *más de un valor nuevo a un solo valor antiguo*.

## Recodificar en distintas variables

Para cambiar los códigos de una variable y, al mismo tiempo, crear una variable nueva con los nuevos códigos (dejando intacta la variable original):

- ▶ Seleccionar la opción **Recodificar > En distintas variables...** del menú **Transformar** para acceder al cuadro de diálogo *Recodificar en distintas variables* (figura 5.10).

**Figura 5.10.** Cuadro de diálogo *Recodificar en distintas variables*.



Este cuadro de diálogo permite crear variables nuevas a partir de los valores de variables ya existentes. Es muy similar al de la figura 5.8. La diferencia está, únicamente, en que aquí es necesario asignar nombre (y, opcionalmente, etiqueta) a las nuevas variables que deseamos crear. Para iniciar la recodificación en variables diferentes:

- ▶ Seleccionar, en la lista de variables, la variable que se desea recodificar y trasladarla, mediante el botón flecha, a la lista **Var. de entrada → Var. de resultado**.
- ▶ En el recuadro **Variable de resultado**, introducir el nombre elegido para la nueva variable (en el cuadro de texto **Nombre**) y, si se desea, una etiqueta descriptiva de hasta 120 caracteres (en el cuadro de texto **Etiqueta**).
- ▶ Pulsar el botón **Cambiar** para activar el nuevo nombre y situarlo en la lista **Var. de entrada → Var. de resultado** junto al nombre de la variable original.

**Si...** La recodificación puede efectuarse de forma condicional, es decir, de forma que sólo afecte a los casos que cumplan determinada condición. Para ello:

- ▶ Pulsar el botón **Si...** para acceder al subcuadro de diálogo *Recodificar en distintas variables: Si los casos* (idéntico al de la figura 5.4). Este subcuadro de diálogo contiene las opciones necesarias para efectuar transformaciones condicionales.

**Valores antiguos y nuevos...** Una vez seleccionadas las variables que se van a recodificar, es necesario definir la recodificación concreta que se desea llevar a cabo. Para ello:

- ▶ Pulsar el botón **Valores antiguos y nuevos...** para acceder al subcuadro de diálogo *Recodificar en distintas variables: Valores antiguos y nuevos* (ver figura 5.11).

**Figura 5.11.** Subcuadro de diálogo *Recodificar en distintas variables: Valores antiguos y nuevos*

Este subcuadro de diálogo es idéntico al de la figura 5.9, pero contiene un elemento adicional en el recuadro **Valor nuevo**:

- **Copiar valores antiguos.** Al marcar esta opción, el valor o valores antiguos seleccionados se mantienen sin cambios en la nueva variable.

Este subcuadro de diálogo también contiene dos elementos adicionales referidos a las variables de cadena:

- ▶ **Las variables de resultado son cadenas.** Si la variable receptora de los nuevos códigos es una variable de cadena, es necesario marcar esta opción y especificar el ancho de la cadena en la caja **Ancho**.

- ▶ **Convertir cadenas numéricas en números.** Esta opción permite transformar una variable de cadena en una variable numérica. Para ello, los valores de la cadena deben ser únicamente números (opcionalmente acompañados de un signo + o −, o de un separador decimal). Si se activa esta opción y la cadena contiene algún carácter no numérico, la cadena se convierte en un valor perdido definido por el sistema.

Los valores antiguos no seleccionados pasan a ser, en la nueva variable, valores perdidos definidos por el sistema. La opción **Todos los demás valores** del recuadro **Valores antiguos** combinada con la opción **Copiar valores antiguos** del recuadro **Valores nuevos** resulta especialmente útil para conseguir que los valores antiguos no seleccionados pasen a formar parte de la nueva variable sin ser convertidos en valores perdidos.

## Categorizar variables

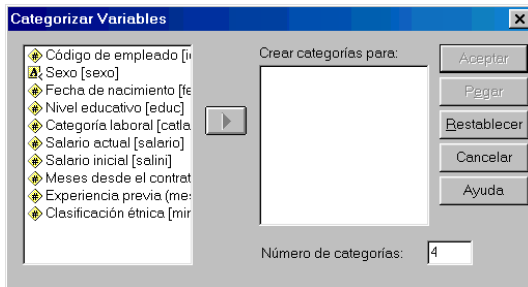
Categorizar variables consiste en crear una variable categórica a partir de una variable cuantitativa continua, tomando como base para la transformación los percentiles de la variable continua. Por ejemplo, la *edad* de los sujetos medida en años y meses, que es una variable cuantitativa continua, puede transformarse en una variable categórica creando cuatro o cinco grupos de edad.

Este tipo de transformación lleva asociada una inevitable pérdida de información, pero puede resultar útil si se tiene interés en construir tablas de contingencia cruzando la variable edad con otras variables categóricas como el sexo o la categoría laboral.; o si se tiene interés en utilizar la edad como variable independiente en, por ejemplo, un análisis de varianza.

Para categorizar una variable:

- ▶ Seleccionar la opción **Categorizar variables...** del menú **Transformar** para acceder al cuadro de diálogo que muestra la tabla 5.12.

**Figura 5.12.** Cuadro de diálogo *Categorizar variables*.



**Crear categorías para.** Las variables continuas que se vayan a categorizar deben trasladarse a esta lista.

**Número de categorías.** En este cuadro de texto debe especificarse el número de categorías que se desea que tenga la nueva variable. Una especificación de, por ejemplo, 2 categorías crea la nueva variable categórica asignando un 1 a los casos con puntuaciones situadas por debajo de la mediana (percentil 50) y un 2 a los casos con puntuaciones situadas por encima de la me-

diana. Y una especificación de, por ejemplo, 4 categorías crea la nueva variable categórica asignando un 1 a los casos situados por debajo del percentil 25, el valor 2 a los casos comprendidos entre el percentil 25 y el 50, el valor 3 a los casos comprendidos entre el percentil 50 y el 75 y el valor 4 a los casos situados por encima del percentil 75. Cada categoría de la nueva variable pasa a tener aproximadamente el mismo número de casos.

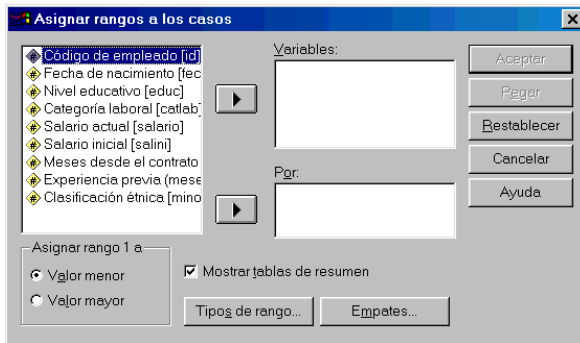
Esta forma de categorizar variables basada en los valores de los percentiles se encuentra también disponible en la opción **Ntiles** del procedimiento **Asignar rangos** que se describe en el siguiente apartado.

## Asignar rangos

Asignar rangos consiste en sustituir los valores originales de una variable numérica por enteros consecutivos de 1 a  $n$ . El SPSS incluye diferentes métodos para llevar a cabo esta sustitución. El SPSS también permite asignar puntuaciones normales y de Savage, y agrupar los casos según el percentil que les corresponde. Para asignar rangos:

- ▣ Seleccionar la opción **Asignar rangos a casos...** del menú **Transformar** para acceder al cuadro de diálogo *Asignar rangos a casos* que muestra la figura 5.13.

**Figura 5.13.** Cuadro de diálogo *Asignar rangos a los casos*.



**Variables.** Para asignar rangos a una variable debe comenzarse seleccionando la variable. Para ello:

- ▣ Marcar la variable en la lista de variables del archivo de datos y trasladarla a la lista **Variables** mediante el botón flecha.

Al pulsar el botón **Aceptar**, el SPSS crea una nueva variable cuyos valores son enteros consecutivos de 1 a  $n$ . La nueva variable recibe de forma automática un nuevo nombre (el nombre de la variable original precedido por una  $r$ ) y una etiqueta. La variable original queda intacta. Si no se marca ninguna otra opción, la asignación de rangos se hace en orden *ascendente* (al valor más pequeño se le asigna un 1) y los empates son sustituidos por la media de los rangos (ver más adelante el apartado **Rangos empatados**).



**Por.** La asignación de rangos puede organizarse por subgrupos. Para ello, es necesario trasladar a esta lista la variable o variables que definen los subgrupos de interés.

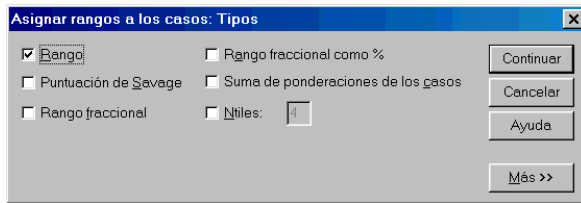
**Asignar rango 1 a.** Las opciones de este recuadro permiten cambiar el orden en el que son asignados los rangos:

- Valor menor.** Asigna los rangos en orden ascendente: al valor más pequeño se le asigna un 1, al valor más pequeño de los restantes se le asigna un 2, etc. Es la opción por defecto.
- Valor mayor.** Asigna los rangos en orden descendente: al valor más grande se le asigna un 1, al valor más grande de los restantes se le asigna un 2, etc
- Mostrar tablas de resumen.** Con esta opción activa (se encuentra activa por defecto), al asignar rangos a una variable el SPSS muestra en el *Visor de resultados* una tabla con el nombre de la variable original, el nombre de la nueva variable y una etiqueta descriptiva incluyendo el tipo de rangos utilizados. El *Visor* no muestra toda esta información si se desactiva esta casilla.

## Tipos de rango

- ▶ Pulsar el botón **Tipos de rango...** (ver figura 5.13) para acceder al subcuadro de diálogo *Asignar rangos a los casos: Tipos* que muestra la figura 5.14.

**Figura 5.13.** Subcuadro de diálogo *Asignar rangos a los casos: Tipos*.



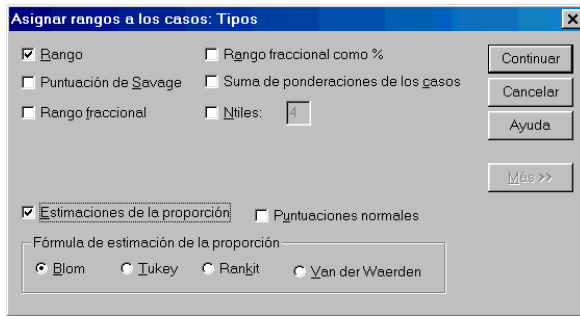
Este subcuadro permite seleccionar diferentes métodos de asignación de rangos. Si se elige más de un método, el SPSS crea una variable diferente con los rangos correspondientes a cada método seleccionado. La etiqueta que el SPSS asigna de forma automática a cada nueva variable informa sobre el método de asignación de rangos utilizado. El SPSS proporciona seis métodos distintos de asignación de rangos:

- Rango.** Asigna enteros consecutivos de 1 a  $n$ . Es la opción por defecto.
- Puntuación de Savage.** Asigna puntuaciones basadas en una distribución exponencial.
- Rango fraccional.** Asigna el resultado de dividir cada rango por el número de casos válidos
- Rango fraccional como %.** Asigna el rango fraccional multiplicado por 100.
- Suma de ponderaciones de casos.** Asigna a cada caso, como único rango, el número de casos válidos. Si utilizamos una variable de agrupación (ver figura 5.13, lista **Por**), a los casos de cada grupo se les asigna como puntuación el número de sujetos del grupo al que pertenecen (por tanto, la puntuación asignada es constante para los casos de un mismo grupo).
- Ntiles.** Esta opción divide la distribución en  $k$  áreas (áreas porcentuales de aproximadamente igual tamaño, basadas en el cálculo de percentiles) y asigna a cada caso, como puntuación, el rango (de 1 a  $k$ ) del área en la que se encuentra. Se obtiene el mismo resultado con la opción **Categorizar variables** del menú **Transformar**.

**Más >>**. Además de varias opciones para asignar rangos a los casos, el cuadro de diálogo *Asignar rangos* contiene también opciones para: 1) obtener estimaciones de la proporción de casos acumulada hasta cada rango y 2) calcular las puntuaciones típicas normales que corresponden a esas proporciones. Para obtener esta información adicional:

- ▶ Pulsar el botón **Más >>**. Esta acción expande el cuadro de diálogo de la figura 5.14 haciéndole mostrar nuevas opciones (ver figura 5.14).

**Figura 5.15.** Subcuadro de diálogo *Asignar rangos a los casos: Tipos* expandido.



- **Estimaciones de la proporción.** Estima la proporción acumulada (el área acumulada de la distribución) que corresponde a cada rango concreto.
- **Puntuaciones normales.** Asigna las puntuaciones típicas  $z$  que corresponden en la curva normal tipificada a las proporciones acumuladas de cada rango. Por ejemplo, a un rango que acumula una proporción de casos de 0,50, se le asigna una puntuación  $z$  de 0; si acumula una proporción de 0,75, se le asigna una puntuación  $z$  de 0,67; etc.

**Fórmula de estimación de la proporción.** La estimación de la proporción acumulada que corresponde a cada rango puede efectuarse utilizando diferentes procedimientos (en todos los casos,  $R_i$  se refiere al rango asignado y  $n$  al número de casos válidos):

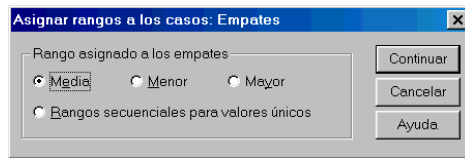
- **Blom.**  $(R_i - 3/8) / (n + 1/4)$ .
- **Tukey.**  $(R_i - 1/3) / (n + 1/3)$ .
- **Rankit.**  $(R_i - 1/2) / n$ .
- **Van der Waerden.**  $R_i / (n + 1)$ .

## Rangos empatados

Cuando existen casos con los mismos valores, es decir, casos *empataados*, a cada caso se le asigna, por defecto, el promedio de los rangos que corresponden a esos casos. Para utilizar un tratamiento distinto de los empates:

- ▶ Pulsar el botón **Empates...** (ver figura 5.13) para acceder al subcuadro de diálogo *Asignar rangos a los casos: Empates* que muestra la figura 5.16.

**Figura 5.16.** Cuadro de diálogo *Asignar rangos a los casos: Empates*.



**Rango asignado a los empates.** El SPSS ofrece cuatro formas distintas de asignar rangos a los empates.

- Media.** Asigna la media de los rangos que corresponden a los casos empatados. Es la opción por defecto.
- Menor.** Asigna el menor de los rangos que corresponden a los casos empatados.
- Mayor.** Asigna el mayor de los rangos que corresponden a los casos empatados.
- Rangos secuenciales para valores únicos.** Asigna rangos de 1 a  $m$ , siendo  $m$  el número de valores distintos. Los casos empatados reciben el mismo rango y cuentan como un único caso al computar el rango siguiente.

El siguiente ejemplo muestra cómo afecta a la asignación de rangos cada una de estas formas de tratar los empates:

<b>Valor</b>	<b>Media</b>	<b>Menor</b>	<b>Mayor</b>	<b>Secuencial</b>
7	1	1	1	1
9	2	2	2	2
11	4,5	3	6	3
11	4,5	3	6	3
11	4,5	3	6	3
11	4,5	3	6	3
14	7	7	7	4
17	8	8	8	5

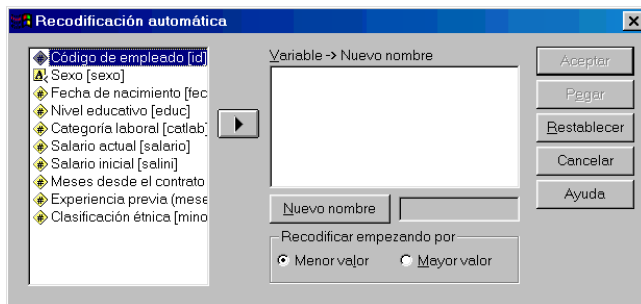
## Recodificación automática

La recodificación automática asigna enteros consecutivos de 1 a  $m$  a los  $m$  valores diferentes de una variable (sea ésta numérica o de cadena). Es, por tanto, de una asignación de rangos equivalente al *método secuencial* de tratamiento de los empates visto en el apartado anterior.

Esta opción resulta particularmente útil si tenemos en cuenta, por ejemplo, que en algunos procedimientos SPSS (como el análisis de varianza) interesa que los niveles de los factores estén codificados como enteros consecutivos. También resulta particularmente útil esta opción para modificar los códigos de las respuestas dadas en una escala de, por ejemplo, actitudes, cuando las puntuaciones altas en unas preguntas indican actitud favorable y en otras preguntas actitud desfavorable y lo que interesa es obtener la puntuación total de la escala como suma de las diferentes preguntas. Para asignar enteros consecutivos:

- ▶ Seleccionar la opción **Recodificación automática...** del menú **Transformar** para acceder al cuadro de diálogo *Recodificación automática* que muestra la figura 5.17.

**Figura 5.17.** Cuadro de diálogo *Recodificación automática*.



**Variable → Nuevo nombre.** Las variables cuyos códigos se van a recodificar en enteros consecutivos deben trasladarse a este recuadro. Para ello, hay que seleccionar, en la lista de variables del archivo de datos, la variable o variables que se van a recodificar y pulsar el botón flecha para trasladarlas a la lista **Variable → Nuevo nombre**.

**Nuevo nombre.** El botón **Aceptar** no está disponible hasta que se da un nombre a la variable receptora de los nuevos códigos. Por tanto, es necesario asociar un nombre nuevo al nombre de la variable original. Para ello:

- ▶ Introducir el nombre deseado en el cuadro de texto situado junto al botón **Nuevo nombre**.
- ▶ Pulsar el botón **Nuevo nombre** para activar el nombre elegido (es decir, para asociarlo al nombre original).

El SPSS crea una variable con los nuevos códigos y el nuevo nombre asignado. La variable original queda intacta. Si la variable original posee etiquetas, a la nueva variable y a los nuevos códigos se les asignan las mismas etiquetas. Si los valores de la variable original no poseen etiquetas, los nuevos códigos adoptan como etiquetas los valores de la variable original.

**Recodificar empezando por.** Las opciones de este recuadro permiten cambiar el orden en el que son asignados los nuevos códigos:

- **Menor valor.** Asigna los códigos en orden ascendente: al valor más pequeño se le asigna un 1, al valor más pequeño de los restantes se le asigna un 2, etc. Es la opción por defecto.
- **Mayor valor.** Asigna los códigos en orden descendente: al valor más grande se le asigna un 1, al valor más grande de los restantes se le asigna un 2, etc.

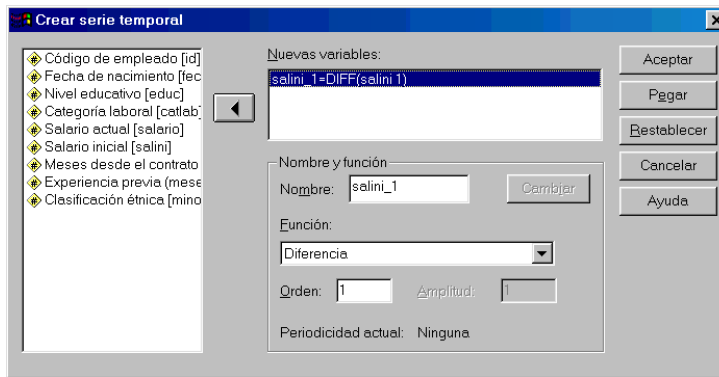
Si la variable es de cadena, los nuevos códigos se asignan de acuerdo con el orden alfabético de las categorías (las mayúsculas preceden a las minúsculas). Los valores perdidos son codificados después de los válidos.

## Crear serie temporal

Esta opción permite generar series temporales a partir de variables existentes. Una serie temporal es una variable cuyos valores progresan a intervalos regulares de tiempo. Estas variables son útiles en muchos de los análisis de series temporales. Para crear una serie temporal:

- ▶ Seleccionar la opción **Crear serie temporal...** del menú **Transformar** para acceder al cuadro de diálogo *Crear serie temporal* que muestra la figura 5.18.

**Figura 5.18.** Cuadro de diálogo *Crear serie temporal*.



**Nuevas Variables.** La variable o variables a partir de las cuales se crearán las nuevas series temporales deben trasladarse a esta lista. Para ello:

- ▶ Seleccionar, en la lista de variables del archivo de datos, la variable o variables a partir de las cuales se van a crear las series temporales (sólo se admiten variables numéricas), y pulsar el botón flecha.

La variable o variables seleccionadas pasan a la lista **Nuevas Variables** con los seis primeros caracteres del nombre original seguidos del guión de subrayado y un número secuencial (la variable *salini*, por ejemplo, ha pasado con el nombre *salini\_1*). Las nuevas variables conservan las etiquetas de las variables originales. El nombre de la nueva variable aparece acompañado de la función que será utilizada para crear la serie temporal (tanto el nuevo nombre como la función los asigna el SPSS por defecto).



**Nombre y función.** El nombre de la nueva variable y la función asignada por defecto pueden cambiarse utilizando estas opciones. Para cambiar el nombre de la nueva variable:

- ▶ Introducir el nuevo nombre en la casilla **Nombre** y pulsar el botón **Cambiar**.

Para cambiar la función:

- ▶ Pinchar con el puntero del ratón en el botón flecha ▼ del menú desplegable **Función** y seleccionar una función de la lista desplegable.

Es posible seleccionar una de las siguientes funciones:

- **Diferencia.** Calcula, para cada caso, la diferencia no estacional entre el valor de ese caso y el valor del caso situado un número determinado de posiciones anteriores. La opción **Orden** permite establecer el número de posiciones que se utilizarán para calcular la diferencia. Dado que al comienzo de la serie se pierden tantos valores como el número de posiciones establecidas en la opción **Orden**, a los primeros casos se les asignan valores perdidos definidos por el sistema. Si, por ejemplo, el *orden* de la diferencia es 3, los primeros 3 casos del archivo tendrán valor perdido en la serie temporal.
- **Diferencia estacional.** Calcula, para cada caso, la diferencia estacional entre el valor de ese caso y el valor del caso situado un número determinado de posiciones anteriores. Ese número de posiciones se basa en la amplitud de un periodo estacional previamente establecido: para calcular diferencias estacionales deben definirse previamente variables *fecha* (ver capítulo 4, apartado *Definir fechas*) que incluyan un componente periódico (como, por ejemplo, los meses del año, o las horas del día, etc.).

La casilla **Orden** recoge el número de periodos estacionales utilizados para calcular la diferencia. El número de casos a los que se asigna un valor perdido al comienzo de la serie es igual a la amplitud del periodo estacional multiplicada por el *orden* establecido. Si, por ejemplo, la periodicidad actual vale 12 (meses del año) y el orden vale 2, se asignará valor perdido a los primeros 24 casos.

- **Media móvil centrada.** Calcula, para cada caso, la media de los valores de los casos que rodean a ese caso (incluido el propio caso). La opción **Amplitud** permite establecer el número de casos que serán utilizados para calcular la media. Si la *amplitud* es par, la media móvil se calcula como el promedio de las dos medias no centradas que corresponden a cada caso. El número de casos a los que se asigna valor perdido al comienzo y al final de la serie es igual a  $k/2$  ( $k = \textit{amplitud}$ ) si la amplitud es par, e igual a  $(k-1)/2$  si la amplitud es impar. Si, por ejemplo, la amplitud vale 5, el número de casos con valor perdido al comienzo y al final de la serie es  $(5-1)/2 = 2$ .

- **Media móvil anterior.** Calcula, para cada caso, la media de los valores de los casos que preceden a ese caso. La opción **Amplitud** permite establecer el número de valores que serán utilizados para calcular la media. El número de casos a los que se asigna valor perdido al comienzo de la serie es igual al valor de la *amplitud*.
- **Medianas móviles.** Calcula, para cada caso, la mediana de los valores de los casos que rodean a ese caso (incluido el propio caso). La opción **Amplitud** permite establecer el número de casos que serán utilizados para calcular la mediana. Si la *amplitud* es par, la mediana se calcula como el promedio de las dos medianas no centradas que corresponden a cada caso. El número de casos a los que se asigna valor perdido al comienzo y al final de la serie es igual a  $k/2$  ( $k = \textit{amplitud}$ ) si la *amplitud* es par, e igual a  $(k-1)/2$  si la *amplitud* es impar. Si, por ejemplo, la *amplitud* vale 5, el número de casos con valor perdido al comienzo y al final de la serie es  $(5-1)/2 = 2$ .
- **Suma acumulada.** Calcula, para cada caso, la suma del valor de ese caso y de todos los que le preceden en la serie.
- **Retardo.** Asigna a cada caso el valor del caso situado un determinado número de posiciones por delante de él. La opción **Orden** permite establecer el número de posiciones en que se basará la asignación. El número de casos a los que se asigna valor perdido al comienzo de la serie es igual al *orden* establecido.
- **Adelanto.** Asigna a cada caso el valor del caso situado un determinado número de posiciones por detrás de él. La opción **Orden** permite establecer el número de posiciones en que se basará la asignación. El número de casos a los que se asigna valor perdido al comienzo de la serie es igual al *orden* establecido.
- **Suavizado.** Calcula, para cada caso, un nuevo valor basado en un largo proceso de suavizado. Este proceso de suavizado comienza con una mediana móvil de *amplitud* 4, que se centra por una mediana móvil de *amplitud* 2. Los valores resultantes se vuelven a suavizar aplicando una mediana móvil de *amplitud* 5 y una mediana móvil de *amplitud* 3, y obteniendo el promedio ponderado de ambas medianas móviles. A continuación se calculan las diferencias (residuos) entre los valores de la serie suavizada y los de la serie original. Después se repite todo el proceso sobre los residuos obtenidos para suavizarlos. Por último, se obtienen las diferencias entre los residuos suavizados y los valores suavizados obtenidos en el primer paso del proceso. A este procedimiento concreto de suavizado se le suele denominar T4253H.

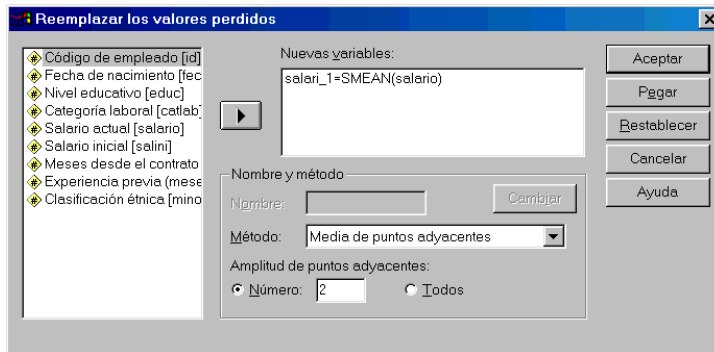
## Reemplazar valores perdidos

Los valores perdidos casi siempre son fuente de problemas en la mayor parte de los procedimientos estadísticos (especialmente en algunos como, por ejemplo, el análisis de series temporales).

Lo ideal es que no existan valores perdidos. Si existen, los problemas que se derivan de su tratamiento pueden resolverse, en parte, sustituyéndolos por alguna estimación del valor que podrían adoptar (aunque este tipo de sustituciones siempre deben realizarse con la máxima cautela). Para reemplazar valores perdidos:

- ▶ Seleccionar la opción **Reemplazar valores perdidos...** del menú **Transformar** para acceder al cuadro de diálogo *Reemplazar valores perdidos* que muestra la figura 5.19.

**Figura 5.19.** Cuadro de diálogo *Reemplazar valores perdidos*.



**Nuevas variables.** Muestra el nombre de la nueva variable junto con el método de estimación que se utilizará para reemplazar los valores perdidos. Para trasladar una variable a esta lista, seleccionar la variable en la lista de variables del archivo de datos y pulsar el botón flecha.

Al seleccionar una variable y trasladarla a la lista **Nuevas variables**, el SPSS le asigna automáticamente un nuevo nombre formado por los primeros 6 caracteres de la variable original, el guión de subrayado y un número secuencial (en el ejemplo, a la variable *salario* le ha asignado el nombre *salari\_1*), y le aplica, por defecto, el método de estimación *media de puntos adyacentes*.

**Nombre y método.** Las opciones de este recuadro permiten cambiar tanto el nombre asignado automáticamente a la nueva variable y el método de estimación utilizado por defecto.

Para cambiar el nombre, basta con introducir el nombre deseado en el cuadro de texto **Nombre** y pulsar el botón **Cambiar**.

El método de estimación puede cambiarse seleccionando cualquiera de los disponibles en el listado obtenido al pulsar el botón de menú desplegable de la opción **Método**:

- **Media de la serie.** Sustituye los valores perdidos por la media aritmética de la serie completa (es decir, por la media aritmética de la variable).
- **Media de puntos adyacentes.** Sustituye cada valor perdido por la media aritmética de sus valores adyacentes válidos.
- **Mediana de puntos adyacentes.** Sustituye cada valor perdido por la mediana de sus valores adyacentes válidos.
- **Interpolación lineal.** Sustituye los valores perdidos utilizando una interpolación lineal basada en el último valor válido anterior al valor perdido y en el primer valor válido posterior al valor perdido. El primer y el último valor de la serie no son reemplazados (en caso de que sean valores perdidos).
- **Tendencia lineal en el punto.** Reemplaza cada valor perdido por los pronósticos obtenidos al efectuar un análisis de regresión lineal utilizando la propia serie como variable dependiente y una variable índice escalada de 1 a  $n$  como variable independiente.

**Amplitud de puntos adyacentes.** En los métodos *Media de puntos adyacentes* y *Mediana de puntos adyacentes* es posible establecer el número de puntos adyacentes en los que deseamos que se base el cálculo de la media o la mediana.

- **Número.** Número de valores válidos, por encima y por debajo del valor perdido, que serán utilizados para el cálculo de la media o la mediana. El número de valores por defecto es 2 (es decir, 2 valores por encima y 2 por debajo). Si la amplitud solicitada es mayor que el número de valores válidos que hay por debajo o por encima de un valor perdido concreto, ese valor perdido no es reemplazado.

- **Todos.** Sustituye los valores perdidos por la media o la mediana de la serie completa. Equivale a seleccionar la opción **Media de la serie**.

< Fin del capítulo 5 >