



Anexo II

*Etiquetado, Recodificación y
Transformación de variables*

Etiquetado, Recodificación y Transformación de variables

1. Etiquetado de variables

Una vez que tenemos incluido en el ordenador todos los datos correspondientes a todas las personas que han sido entrevistadas y de todas las variables de cada una de las preguntas del mismo, pasamos a etiquetar cada una de esas variables, definiendo tanto el tipo de variable como las categorías que posee.

Ya hemos comentado en el anexo explicativo anterior que las preguntas de un cuestionario poseen una o más de una variable, dependiendo del tipo de pregunta. Y también sabemos que las variables de una misma pregunta las distinguiremos colocando una letra detrás del nombre de la variable. Con la secuencia de pasos que exponemos para etiquetar las variables conseguimos sustituir la serie de letras y números con los que inicialmente hemos introducido las variables por un nombre o “etiqueta” más literal y que en definitiva nos permitirá saber exactamente y en todo momento (fundamentalmente en la ejecución de gráficos e informes estadísticos) con qué variable conceptual estamos operando.

1^{er} paso: Para empezar a etiquetar una variable, cliquearemos dos veces sobre el nombre de la variable que aparece en la parte superior de la matriz del SPSS (en fondo gris). Una vez que lo hagamos, nos aparecerá la ventana Definir variable en donde empezaremos a especificar todas las características de dicha variable (figura 1). En la parte superior, figura el Nombre de la variable.

2^o paso: En la parte central se encuentra el botón de comando Tipo de variables. Cliqueando sobre él accedemos al subcuadro de diálogo donde podremos especificar, y valga la redundancia, el tipo de datos que van asociados a esa variable. Aunque normalmente los datos son numéricos, ésta es la opción que aparece por defecto, este subcuadro ofrece otros “tipos” debiendo elegir aquel que mejor represente el tipo de valor asociado con la variable que estemos etiquetando. En este subcuadro

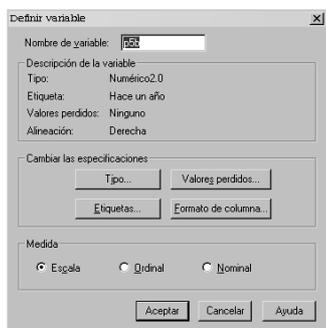


Figura 1

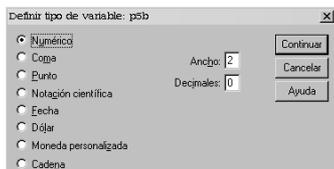


Figura 2

también especificaremos los decimales, si los llevara, y el ancho de la casilla (el número de caracteres) (figura 2).

3^{er} paso: Una vez determinado el tipo de dato asociado a la variable se especifican las Etiquetas de las variables con las que vamos a trabajar. Para acceder a su correspondiente subcuadro de diálogo deberemos clicar sobre el botón de comando que le da nombre, también situado en la parte central del cuadro de diálogo principal (ver figura 1). Este subcuadro nos permite:

- Escribir la etiqueta de la variable, o lo que es lo mismo, el nombre literal de la misma. En Etiqueta de la variable escribimos el nombre propiamente de la variable. Si al crear nuestro archivo de datos colocábamos la numeración de la pregunta que aparecía en el cuestionario donde se encontraba la variable, para localizarla fácilmente, ahora, escribimos el nombre de la variable de manera más formal, intentando dar con el nombre la mayor información posible. Recordemos que la finalidad del etiquetado no es otra que la de facilitar la interpretación y análisis de los resultados objetivo difícil de conseguir si sólo sabemos los números de las preguntas que relacionamos.
- Escribir las etiquetas de valor. En Valor indicamos el valor de cada una de las posibles contestaciones a esa pregunta y en Etiquetas de valor la etiqueta descriptiva asociada a cada uno de los valores. Por ejemplo, la variable SEXO consta de dos valores: hombre y mujer. Para introducir el primero de ellos, en Valor escribiríamos 1; y en Etiqueta de valor el descriptivo HOMBRE. Para introducir el segundo valor de la variable sexo deberemos previamente pulsar sobre el botón AÑADIR. Volveremos a valor y escribiremos el valor 2; y en Etiqueta de valor escribiremos el descriptivo MUJER (figura 3). Una vez finalizada la asignación de etiquetas de valor al valor de los datos terminaremos pulsando el botón Continuar.

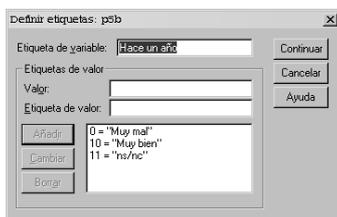


Figura 3

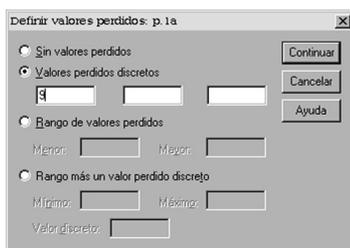


Figura 4

4^o paso. Por último, se determina si hay Valores perdidos o no. Los valores perdidos son: contestaciones de los encuestados al no sabe (NS) o al no contesta (NC); o bien, son casillas que se han quedado en blanco. El SPSS ofrece esta opción para poder evitar el efecto de estas contestaciones a cualquiera de los análisis que realicemos sobre la matriz de datos ya que pueden

distorsionar la realidad. Se colocara el valor donde se encuentre esa posible contestación, en nuestro ejemplo, el valor 9 (figura 4). Algunas consideraciones sobre el tratamiento de la no respuesta se exponen en el último anexo explicativo que figura a continuación.

1.1. Etiquetado de variables con el SPSS 10.

En la última versión del SPSS (10.) el etiquetado sigue la secuencia descrita con la salvedad de que la opción Definir variable ya no aparece en el menú Datos sino que aparece automáticamente una vez que abrimos el programa (ya sea con Nuevo o con Abrir un archivo ya existente).

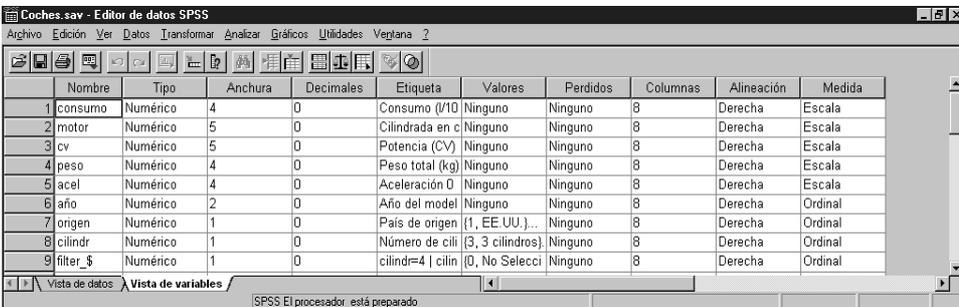


Figura 5

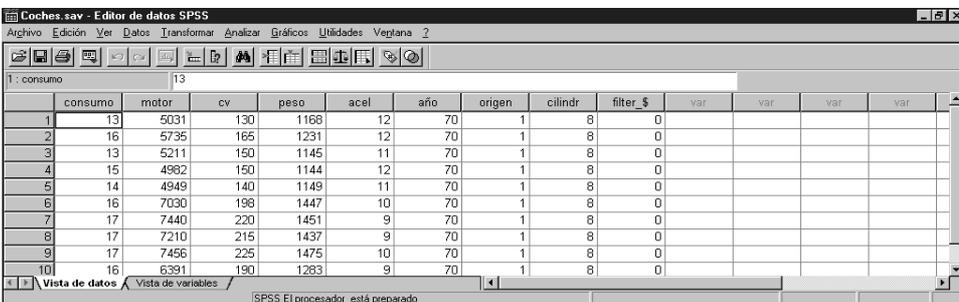


Figura 6

Al abrir o iniciar cualquier archivo aparece detrás de la pantalla con la matriz o estructura de datos (el editor del SPSS), la pantalla para definir las variables. Para acceder a una u otra deberemos clicar la “pestaña” correspondiente: la Vista de datos,

para ver la matriz de datos tabulada; y la Vista de Variables, para ver las características de cada una de las variables (ambas se encuentran en la esquina inferior izquierda) (figura 5).

Al pulsar sobre Vista de variables aparecen en las filas las variables y en las columnas las características que habrá que indicar sobre cada una de ellas, a saber: Nombre, Tipo, Anchura, Decimales; Etiqueta, Valores, Perdidos, Columnas, Alienación y Medida (figura 6). Al pulsar sobre cada una de ellas o bien aparece una ventana desplegable o bien aparece el correspondiente cuadro de diálogo que no varía de lo descrito hasta ahora.

2. Recodificación de variables

Algunos de los ejemplos que hemos recogido en este libro ejecutaban el análisis correspondiente a partir de alguna de sus variables recodificadas. El objetivo de la recodificación de una variable no es otro que el de reasignar los valores de la variable o agrupar rangos de valores existentes en nuevos valores. Así por ejemplo, nos podría interesar, agrupar a los municipios objetos de estudios en municipio de pequeño tamaño, municipios medios, grandes ciudades y áreas metropolitanas; o bien agrupar en grupos de edad a la población encuestada. Con ello nos evitaríamos trabajar con tantas categorías como número de habitantes o como edades tuviéramos. Pueden recodificarse múltiples variables siempre y cuando éstas sean de igual tipo. Por su parte sólo se recodifican las variables numéricas y de cadena.

1^{er} paso: El cuadro de diálogo para recodificar se encuentra en el menú de Transformar, por ello la ruta que nos lleva a él es la que sigue, Transformar: Recodificar: En distintas variables (figura 1). Elegimos la opción en Distintas variables pues de esta manera añadimos a la variable fuente una nueva variable recodificada. Así, no perdemos la variable original.



Figura 1



Figura 2

2^o paso: Una vez en este cuadro de diálogo deberemos especificar qué variables queremos recodificar seleccionándolas del listado que aparece en la ventana de la izquierda y colocándolas en la ventana Variables numéricas. Nosotros vamos a recodificar la variable d3 ESTADO CIVIL. Esta es la recodificación que aplicamos en el capítulo que dedicamos al análisis de tablas de contingencia. En el mismo cuadro de diálogo daremos un Nombre a la Variable de resultado que se genera una vez concluido el proceso de recodificación (figura 2). En nuestro

caso la nombramos como rd3 y, a continuación, pulsamos Cambiar. La variable numérica junto a la variable de resultado (ésta sustituye al signo interrogante que aparece inicialmente) aparecerán, en este momento, en la ventana correspondiente.

3^{er} paso: Una vez seleccionada la/s variable/s a recodificar hay que especificar las nuevas categorías. Para ello cliqueamos sobre el botón de comando Valores antiguos y nuevos... que aparece en la parte inferior del cuadro de diálogo principal de recodificar. El subcuadro de diálogo Valores antiguos y nuevos aparece dividido en dos partes: la parte izquierda la dedica a especificar los valores antiguos o asociados al etiquetado inicial de la variable objeto de recodificación; mientras que la parte derecha se destina a concretar los valores nuevos que adoptará la variable una vez que hayamos concluido el proceso.



Figura 3



Figura 4

- La primera categoría asumía el Valor antiguo 1; esta categoría permanece igual por lo que en Valor nuevo volveremos a poner 1. Este proceso, una vez clicado sobre añadir, se repite con el Valor antiguo 2 que también permanece inalterable, primeras conservan el etiquetado inicial (figura 3).
- En tercer lugar, las categorías con Valores antiguos 3, 4 y 5 van a pasar a adoptar el Valor nuevo 3. Dado que el nuevo valor va estar compuesto por un rango de valores antiguos debemos especificar en la mitad izquierda del cuadro dicho rango. Para ello seleccionaremos el botón de Rango (automáticamente se desactiva el de Valor) especificando en las casillas situadas abajo 3 hasta 5 (figura 4). Pulsamos Añadir.
- La sexta categoría cuyo Valor antiguo era el de 6 pasa ahora a adoptar el Valor nuevo 4. Pulsamos a Añadir.
- Y para finalizar, el Antiguo valor 9 mantiene en Valor nuevo el mismo. Pulsamos a Añadir. En la ventana de la derecha aparece la recodificación especificada en este tercer paso. Una vez finalizado el proceso cliqueamos sobre Continuar y ya en el cuadro de diálogo principal Aceptamos el proceso descrito.

4^o paso: Por último, y con el objetivo de poder identificar con facilidad en análisis y gráficos las nuevas categorías, necesi-

tamos etiquetarlas. Para ello, seleccionamos en el Editor de datos la variable creada, rd3 (ésta ocupará la última posición en la matriz de datos) y en Datos seleccionamos la opción Definir variable. Una vez en el cuadro de diálogo principal de definir variables cliquemos sobre: (1) el botón de comando Tipo y en decimales pondremos 0; y (2) sobre el botón de comando Etiquetas y allí procederemos del modo ya descrito en el primer apartado de este anexo explicativo.

3. Transformación de variables

En el capítulo dedicado al análisis exploratorio ya expusimos en qué consistía el proceso de transformación de variables. En este punto recuperamos aquella exposición pues pensamos que éste es un aspecto que trasciende del citado capítulo. Este es un aspecto al que se recurre independientemente de que si se está efectuando un análisis exploratorio o no. Para corroborar lo expuesto ver el capítulo dedicado al análisis de varianza.

Una variable se transforma con la intención de que su distribución sea si no normal muy próxima a ella. Cabe recordar que este es uno de los principios más importantes en los contrastes y análisis paramétricos. El favorecer la normalidad en la distribución de una variable no es sólo una característica deseable en si misma sino que, además, propicia la linealidad en su relación con el resto de las variables.

El SPSS contempla un amplio abanico de posibilidades a la hora de transformar las variables. Para optar por una u otra transformación podemos consultar dos resultados que el análisis exploratorio nos ofrece. En primer lugar, debemos recordar que el test K-S nos indicaba si era o no necesaria la transformación; y en segundo lugar, los gráficos de “tallos y hojas” y/o “cajas con bigotes”, al representar la dirección de la asimetría, nos indica el tipo de transformación más indicada ya que éstas se clasifican atendiendo a la asimetría que manifiestan las distribuciones de las variables; esto es, según sean asimétricas positivas (mediana más cerca de la parte inferior de la caja, concentración de casos en los valores inferiores de la distribución y cola extendida hacia los valores grandes) o, asimétricas negativas (mediana más cerca de la parte superior de la caja, casos concentrados en los valores altos y cola alargada hacia los valores inferiores de la distribución). Junto a las representaciones gráficas puede resultar

de utilidad consultar los estadísticos que hacen referencia a la forma de las distribuciones.

Tukey ofrece lo que él llama la “Escalera de las transformaciones”, donde muestra el tipo de transformación recomendada según sea la intensidad de la asimetría o la dirección en la que van los casos extremos.

- Si la distribución es asimétrica positivas (x^2 , x^3 , anti-log x) (la mediana más cerca de la parte inferior de la caja, concentración de los casos en los valores inferiores de la distribución y cola extendida hacia los valores grandes): elevar al cuadrado o al cubo, así como calcular antilogaritmos, tiene el efecto de corregir la asimetría (más fuerte la corrección cuanto mayor sea la exponenciación).
- Para corregir distribuciones asimétricas negativas ($\log x$, $-1/x$, $-1/x^2$) (Mediana más cerca de la parte superior de la caja, concentración de los casos en los valores altos de la distribución y la cola alargada hacia los valores inferiores de la distribución): es conveniente utilizar raíces cuadradas, logaritmos, etc.

Las transformaciones más comúnmente utilizadas son: “logaritmo (ln)”, “raíz cuadrada (SQRT)” y “potencias (**2)”.

Para acceder al Cuadro de Diálogo seguir la (figura 1) Transformar: Calcula. Los pasos del proceso son los que siguen:

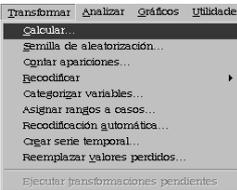


Figura 1

1er paso: En Variable de destino (figura 2) indicar el nombre de la nueva variable que va a crearse a partir de la transformación que señalemos.

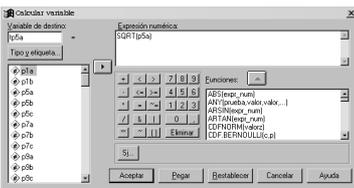


Figura 2

2º paso: A continuación, seleccionar del Cuadro de Funciones la operación matemática adecuada según el tipo de asimetría que tengamos y llevarla al Cuadro Expresión numérica.

3er paso: Por último, y para completar la expresión numérica anterior, indicar la variable original, situada en el Cuadro Tipo y etiqueta, que quiere transformarse mediante esa función llevándola a este mismo cuadro. Finalizado el proceso se obtiene una

nueva variable, ocupará la última columna en la matriz de datos, fruto de un proceso de transformación. De este modo, la variable original, la que se decidió transformar, guarda sus valores originales.

La escalera de transformaciones de Tukey es una buena guía orientativa sobre el tipo de transformación a realizar. Pese a ello, no deja de ser una “guía” ya que a efectos prácticos en raras ocasiones conseguimos la distribución normal anhelada con la primera transformación. Por ello, y una vez obtenida la primera transformación, a la distribución obtenida, la someteremos a los test de normalidad ya descritos. Si con el proceso no llegamos a los resultados requeridos, volveremos a repetir la operación y secuencia de pasos descritos en esta última sección. Sin embargo, conviene advertir que con variables que muestran una distribución próxima a la normalidad la aplicación de diferentes transformaciones provoca hacerlas más asimétricas. Por ello, y ante estas circunstancias, lo más conveniente es seguir trabajando con la variable original.