
Práctica de Estadística con SPSS

Manipulación de datos y ficheros

Ejercicio

Se ha diseñado un ensayo clínico aleatorizado, doble-ciego y controlado con placebo, para estudiar el efecto de dos alternativas terapéuticas en el control de la hipertensión arterial. Se han reclutado 100 pacientes hipertensos y estos han sido distribuidos aleatoriamente en tres grupos de tratamiento. A uno de los grupos (control) se le administró un placebo, a otro grupo se le administró un inhibidor de la enzima convertidora de la angiotensina (IECA) y al otro un tratamiento combinado de un diurético y un Antagonista del Calcio. Las variables respuesta final fueron las presiones arteriales sistólica y diastólica.

Los datos con las claves de aleatorización han sido introducidos en una base de datos que reside en la central de aleatorización, mientras que los datos clínicos han sido archivados en dos archivos distintos, uno para cada uno de los dos centros participantes en el estudio.

Las variables almacenadas en estos archivos clínicos son las siguientes:

- CLAVE Clave de aleatorización
- NOMBRE Iniciales del paciente
- F_NACIM Fecha de Nacimiento
- F_INCLUS Fecha de inclusión
- SEXO Sexo (0: Hombre 1: Mujer)
- ALTURA Altura en cm.
- PESO Peso en Kg.
- PAD_INI Presión diastólica basal (inicial)
- PAD_FIN Presión diastólica final
- PAS_INI Presión sistólica basal (inicial)
- PAS_FIN Presión sistólica final

El archivo de claves de aleatorización contiene sólo dos variables.

- CLAVE Clave de aleatorización
- FARMACO Fármaco administrado (0: Placebo, 1: IECA, 2:Ca Antagonista + diurético)

Se pide:

1. Leer los datos del centro con 10 pacientes, incluidos en el archivo Hipertensos HA.xls, este hospital trabaja con la hoja de cálculo Excel.

Indicación

Desplegar el menú **Archivo->Abrir->Datos**, e ir hasta el directorio que contiene el archivo Hipertensos HA.xls, escogiendo como tipo de archivo el formato de Excel. Una vez tenemos delante el nombre del archivo, para abrirlo será suficiente con un doble clic, pero teniendo en cuenta que hay que activar la opción **Leer nombre de variables** si queremos que utilice la primera fila del archivo de Excel para dar nombre a las variables en el archivo de datos de SPSS.

- Añadir a estos, los datos de los pacientes 11 al 100, incluidos en el archivo Hipertensos HB y guardar los datos como Hipertensos totales.

Indicación

- Teniendo en cuenta que lo que pretendemos es añadir nuevos casos a un archivo de datos ya abierto, el menú a utilizar es **Datos->Fundir archivos->Añadir casos**. Después activamos la opción **Un archivo de datos de SPSS externo**, y con el botón **Examinar** accedemos hasta la carpeta que contenga el archivo Hipertensos HB. Una vez seleccionado, utilizamos el botón **Continuar**, y posteriormente el botón **Aceptar** en el siguiente cuadro de diálogo que aparece.
- Una vez generado el archivo de datos, para guardarlo podemos utilizar el menú **Archivo->Guardar como**.

- Fusionar los datos clínicos con las claves de aleatorización. El fichero con las clave se denomina Claves aleatorizacion. Grabar el archivo resultante con el nombre Hipertensos Datos Claves

Indicación

De forma similar al apartado anterior pero teniendo en cuenta que ahora lo que pretendemos es añadir variables. Para ello: **Datos->Fundir archivos->Añadir variables**. Para acceder después hasta el archivo Claves aleatorizacion, y posteriormente vamos aceptando en todos los cuadros de diálogo que aparecen. Una vez generado, guardamos el nuevo archivo de datos y claves: menú **Archivo->Guardar como**.

- Crear un archivo para cada uno de los grupos de tratamiento. Denominar a estos archivos Hipertensos placebo, Hipertensos IECA e Hipertensos Ca respectivamente.

Indicación

Se podría lograr el mismo resultado de múltiples maneras. Por ejemplo:

- Segmentando el archivo mediante el menú **Datos->Segmentar archivo**, y **Organizar los resultados por grupos** basados en la variable **farmaco**. Una vez segmentado el archivo, podemos marcar los casos correspondientes a cada uno de los fármacos y hacer un cortar y pegar en un nuevo archivo específico para cada fármaco.
- También podemos generar los nuevos archivos utilizando el sistema de filtros de SPSS. Para ello el menú a utilizar es **Datos->Seleccionar Casos**, con la opción **Si se satisface la condición**, botón **Si**, y entramos en un cuadro de diálogo en el que damos forma a la condición, que en primera instancia será **farmaco=1**, para volver al cuadro de diálogo anterior y escoger la opción **Eliminar casos no seleccionados**. Con ello nos quedará un archivo de datos que tan sólo contiene los casos correspondientes al primer fármaco, y podemos hacer un **Archivo->Guardar como** para ponerle el nombre adecuado. Igualmente, repetiríamos el proceso para los otros dos fármacos.

- Calcular, para cada paciente, la edad en años el día de la incorporación al estudio (redondeando al entero más próximo). Denominar la nueva variable **edad** y etiquetarla correspondientemente.

Indicación

Internamente SPSS trabaja con las variables en formato fecha almacenando el número de segundos transcurridos desde el comienzo del Calendario Gregoriano en el año 1582 hasta la fecha que introducimos. Por lo tanto, si restamos dos variables en formato fecha, no nos da el número de años transcurridos entre una y otra, sino el número de segundos. Por ello, la nueva variable obtenida como resultado de la resta hay que dividirla entre el número de segundos que tiene un año a razón de 3600 segundos la hora, 24 horas el día, y 365,25 días el año, aproximadamente. Teniendo en cuenta lo anterior, el proceso a utilizar es: Menú **Transformar->Calcular variable**, opción **Variable de destino**, y damos nombre a la variable **edad**. Como expresión numérica para su cálculo introducimos:

$$(F_INCLUS - F_NACIM)/(365,25 \times 24 \times 3600)$$

- Recodificar dicha edad de forma que la nueva variable, de nombre **grupoedad**, tome los siguientes valores y etiquetas de valor:

Edad en años	Grupo edad	Etiqueta
Menores de 37	1	Menores de 37
De 37 a 44	2	De 37 a 44
De 45 a 52	3	De 45 a 52
Mayores de 52	4	Mayores de 52

Indicación

Para recodificar una variable, se utiliza el proceso ya explicado en la práctica de Introducción a SPSS, con el menú **Transformar->Recodificar en distintas variables**, escogiendo la **Variable de entrada**, dando nombre a la **Variable de resultado**, utilizando el botón **Cambiar**, y posteriormente el botón **Valores antiguos y nuevos** para delimitar las categorías de la nueva variable.

7. Calcular, para cada paciente, el índice de masa corporal (se obtiene dividiendo el peso, expresado en kg, entre la altura, expresada en m, elevada al cuadrado) y almacenar el resultado en la variable **masacorp**.

Indicación

Con el menú **Transformar->Calcular variable**, escogiendo como variable de destino **masacorp**, y como expresión numérica la indicada en el enunciado.

8. Recodificar dicho índice de masa corporal, de forma que la nueva variable, de nombre **obesidad**, tome los siguientes valores y etiquetas de valor según el sexo del paciente.

Sexo	Masacorp	Obesidad	Etiqueta
0	Menor de 21	1	Desnutrido
	De 21,01 a 26,94	2	Normal
	De 26,95 a 32,94	3	Sobrepeso
	De 32,95 a 43,94	4	Obeso
	Mayor que 43,95	5	Muy obeso
1	Menor de 19	1	Desnutrido
	De 19,01 a 24,94	2	Normal
	De 24,95 a 29,94	3	Sobrepeso
	De 29,95 a 39,94	4	Obeso
	Mayor que 39,95	5	Muy obeso

Indicación

Se trata de un problema de recodificación, muy parecido al explicado en la indicación del punto 6, con la única novedad de que ahora hay que hacer una doble recodificación: por un lado para hombres y por el otro para mujeres. Para ello, después de escoger las variables de entrada y de resultado, haciendo uso del botón condicional **Si** podemos escoger únicamente los casos que cumplen la condición impuesta por la variable **Sexo**; es decir, con la condición **Sexo=0**, recodificamos en primera instancia el índice de masa corporal de los hombres, y con **Sexo=1** recodificamos el de las mujeres.