

PRÁCTICA: ESTADÍSTICA DESCRIPTIVA CON SPSS¹

1. Introducción

El programa estadístico SPSS está organizado en dos bloques: el *editor de datos* y el *visor de resultados*. En los dos apartados siguientes vamos a describir las principales características de ambos.

1.1. El editor de datos

Tiene una estructura similar a la de una hoja de cálculo y se utiliza para introducir los datos que se quieren analizar. Los datos se introducen dentro de variables. Cada variable aparece como una columna en la pantalla. En la fila superior del editor de datos aparece un menú con diferentes opciones. Cada una de ellas da acceso a un submenú. A continuación se indican las principales opciones junto con una breve descripción de la utilidad de cada una de ellas. Puede observarse que algunas de ellas (**Archivo**, **Edición**, **Ver**, **Utilidades**, **Ventana** y **?**) son comunes a muchos programas basados en Windows, mientras que **Datos**, **Gráficos**, **Analizar** y **Transformar** son específicas del SPSS.

El contenido genérico de las opciones más utilizadas es el siguiente:

- Archivo** Manejo de ficheros. Importación y exportación de ficheros de datos.
- Edición** Seleccionar, “cortar y pegar” textos, etc.
- Ver** Apariencia de la pantalla. Visualizar datos ó variables.
- Datos** Definición y manejo de variables. Transponer, Ordenar, Seleccionar datos.
- Transformar** Operaciones con los datos de las variables, funciones aplicadas a los datos, generación de números aleatorios, crear series temporales, etc.
- Analizar** Menú de métodos estadísticos para ser aplicados sobre los datos incluidos en las variables. Estadísticos descriptivos, Regresión, Series temporales, etc.
- Gráficos** Diferentes representaciones gráficas (diagramas de barras, de sectores, histogramas, diagramas de caja, etc.) de los datos.
- Utilidades** Para obtener información sobre variables y archivos, definir conjuntos de variables y utilizar sólo las variables seleccionadas en esos conjuntos.
- ?** Ayuda en línea.

1.2. El visor de resultados

Es la ventana en la que se muestran los resultados. Se divide en dos paneles: el panel izquierdo muestra un esquema de los contenidos mientras que el panel derecho contiene tablas estadísticas, gráficos y resultados de nuestros análisis. Los resultados se pueden editar de manera que es factible:

- Mostrar u ocultar tablas y gráficos seleccionados. Modificar columnas o filas de una tabla. Modificar los tipos de puntos, el color, el relleno y otras características de un gráfico en el *Editor de gráficos*. Cuando se termine de modificar una tabla o un gráfico, se debe **cerrar** el editor correspondiente.
- Mover elementos entre el visor y otras aplicaciones (por ejemplo, exportar tablas y gráficos a otros programas).

¹Para escribir estas notas se ha utilizado la versión SPSS 19

2. ¿Cómo introducir datos en SPSS?

Hay dos posibilidades en función de si nosotros tenemos que introducir los datos o si ya disponemos de un fichero que los contiene:

1. Introducir los datos directamente en el editor de datos de SPSS. Si se van a volver a utilizar los datos en sesiones futuras, se utiliza **Archivo** → **Guardar**, o **Archivo** → **Guardar como...** para guardar el fichero .sav resultante.
2. Si se tienen los datos en un fichero de texto o fueron introducidos y salvados previamente con SPSS u otro programa, se elige **Archivo** → **Abrir** → **Datos**. En el cuadro de diálogo que aparece se selecciona la extensión del archivo que queremos abrir (.xls, .dat, .txt, etc.). A continuación sólo hay que seguir con cuidado las opciones del asistente para la importación de datos.

Ejemplo: La tasa metabólica, es decir, la tasa a la que el cuerpo consume energía, es relevante en los estudios sobre ganancia de peso y dietética. El fichero **metabolismo.sav** contiene datos sobre masa corporal magra (en kg) y tasa metabólica (en cal por cada 24 horas) de 12 mujeres y 7 hombres que participaron en un estudio de dietética. Una vez hemos puesto el fichero en una carpeta conveniente, para poder trabajar con los datos debemos seguir los siguientes pasos:

- **Archivo** → **Abrir** → **Datos**.
- En el cuadro de diálogo que aparece, elegir la carpeta en la que se encuentra el archivo y la extensión (en nuestro caso **.sav**, que es la opción por defecto) y seleccionar el archivo **metabolismo.sav**.

3. Gráficos de barras y de sectores

Para construir gráficos de barras y de sectores, hay que seleccionar la opción

Gráficos → **Cuadros de diálogo antiguos** → **Barras**, o la opción
Gráficos → **Cuadros de diálogo antiguos** → **Sectores**.

A continuación tenemos que elegir el tipo de diagrama e indicar como se presentan los datos. Para ello, tenemos en cuenta las siguientes observaciones:

- Las opciones **Agrupado** y **Apilado** sólo tienen sentido cuando queremos construir el diagrama de barras para más de una variable.
- Si tenemos una variable con los resultados del experimento (cada dato es una categoría) elegimos la opción **Resúmenes para grupos de casos**.
- Si tenemos una variable con los nombres de las clases y otra con las frecuencias con las que aparece cada clase, elegimos la opción **Variables individuales de los casos**.

Ejemplo (continuación): Construye un diagrama de barras de la variable **Sexo**:

- **Gráficos** → **Cuadros de diálogo antiguos** → **Barras**.
- Activar las opciones **Simple** y **Resúmenes para grupos de casos** → **Definir**.
- Podemos especificar el tipo de diagrama que deseamos: número de casos (frecuencias absolutas), casos acumulados, % de casos (frecuencias relativas), o % de casos acumulados. Por ejemplo, seleccionemos % de casos. Pasamos la variable **Sexo** a la opción **Eje de categorías** → **Aceptar**.

4. Histogramas

Existen varias posibilidades para construir histogramas, veamos una sencilla. Supongamos que queremos construir un histograma para la variable **Masa**:

- **Gráficos** \leftrightarrow **Cuadros de diálogo antiguos** \leftrightarrow **Histograma...**.
- Pasar la variable **Masa** a **Variable** \leftrightarrow **Aceptar**.

Si deseamos representar un histograma para los hombres y otro para las mujeres, pasamos también la variable **Sexo** a **Panel por Filas** o **Panel por Columnas** antes de **Aceptar**.

Para modificar el aspecto del gráfico (colores, aspecto de los ejes, etc.), tenemos que entrar en el editor de gráficos, haciendo doble clic sobre el gráfico.

Haciendo doble click en un histograma y seleccionando la opción **Mostrar curva de distribución** del menú **Elementos**, se añade al histograma la curva de la distribución normal.

5. Estadísticos descriptivos

Para obtener los estadísticos descriptivos (media, mediana, moda, varianza, cuartiles, etc.) una posible secuencia de pasos es la siguiente:

- **Analizar** \leftrightarrow **Estadísticos descriptivos** \leftrightarrow **Explorar... .**
- Pasar las variables que se deseen estudiar a **Lista de dependientes**.
- Si se desean calcular los estadísticos descriptivos por separado para cada grupo definido por otra variable (por ejemplo, la variable **sexo**) se debe pasar esta variable a **Lista de factores**.
- **Aceptar**.

Ejercicio. Con los datos del ejemplo, calcula las principales medidas descriptivas para la tasa metabólica y la masa corporal magra. Representa también dos histogramas de la tasa metabólica (para los hombres y las mujeres por separado).

6. Diagrama de cajas

Vamos a comparar la tasa metabólica (variable cuantitativa continua) de los hombres y las mujeres (variable cualitativa) mediante el uso de diagramas de cajas. Los pasos a seguir son:

1. Se selecciona **Gráficos** \leftrightarrow **Cuadros de diálogo antiguos** \leftrightarrow **Diagramas de cajas**.
2. Se eligen las opciones **simple** y **luego Resúmenes** para grupos de casos y, a continuación, se pulsa **Definir**.
3. En el cuadro **Variable** se introduce la variable que contiene las tasas metabólicas. En el cuadro **Eje de categorías** se introduce la variable que contiene el sexo. Finalmente, se pulsa **Aceptar**. Una vez obtenido el gráfico, compara la posición, la dispersión y la asimetría de la distribución de la tasa metabólica en los hombres y en las mujeres.

7. Diagrama de dispersión, correlación y recta de mínimos cuadrados

Para estudiar gráficamente el grado de asociación existente entre dos variables elegimos la opción:

Gráficos \leftrightarrow **Cuadros de diálogo antiguos** \leftrightarrow **Dispersión/Puntos**

1. Se elige la opción **Simple** y, a continuación, se pulsa **Definir**

2. Como variable *Y* debemos elegir la variable *dependiente*, la que queremos explicar. Por ejemplo, la tasa metabólica. Como variable *X* elegimos la variable *explicativa*. Por ejemplo, la masa corporal magra.
3. Si queremos que en el gráfico se utilice un color diferente según el sexo, debemos introducir la variable *Sexo* en el cuadro *Establecer marcas por*.
4. Finalmente, se pulsa **Aceptar**.

Una vez obtenido el gráfico, ¿puede decirse que la relación entre la masa corporal magra y la tasa metabólica es lineal? ¿Es la asociación positiva o negativa?

Si queremos añadir al gráfico la representación gráfica de la recta de mínimos cuadrados hacemos doble click en el gráfico y seleccionamos, en el menú *Elementos*, la opción *Línea de ajuste total*.

Los valores numéricos de la pendiente y el término independiente de la recta se pueden calcular utilizando la siguiente opción:

Analizar \leftrightarrow Regresión \leftrightarrow Lineales...

En el correspondiente cuadro de diálogo debemos elegir las mismas variables dependiente e independiente de antes.

Para calcular la correlación entre las dos variables, elegimos

Analizar \leftrightarrow Correlaciones \leftrightarrow Bivariadas

y pasamos al cuadro **Variables** todas las variables entre las que queremos calcular correlaciones. ¿Cuál es el valor de la correlación entre la tasa metabólica y la masa corporal magra?

8. Transformación de variables

Para crear una nueva variable como resultado de aplicar una transformación a alguna de las ya existentes, se procede de la forma siguiente:

- **Transformar \leftrightarrow Calcular variable...**.
- En **variable de destino** se escribe el nombre de la nueva variable que vamos a crear. En los cuadros **Grupo de Funciones** y **Funciones y variables especiales**, se elige la transformación deseada. Por ejemplo, si queremos calcular el logaritmo neperiano de una variable, seleccionamos **Aritméticas** y **LN** respectivamente.
- Aparece en el recuadro de arriba un signo de interrogación que tenemos que sustituir por la variable que queremos transformar.

Ejercicio. Con los datos del ejemplo, construye histogramas de los logaritmos de las tasas metabólicas de los hombres y las mujeres. Calcula la recta de mínimos cuadrados que relaciona el logaritmo de la masa corporal magra con el logaritmo de la tasa metabólica.

9. Ejercicio 8 del tema 1

En 1778, H. Cavendish realizó una serie de 29 experimentos con objeto de medir la densidad de la tierra. Sus resultados, tomando como unidad la densidad del agua, fueron:

5'50 5'61 4'88 5'07 5'26 5'55 5'36 5'29 5'58 5'65
 5'57 5'53 5'62 5'29 5'44 5'34 5'79 5'10 5'27 5'39
 5'42 5'47 5'63 5'34 5'46 5'30 5'75 5'68 5'85

a) Representa los datos por medio de un diagrama de tallos y hojas.

- b) Representa los datos por medio de un diagrama de caja y bigotes.
- c) Halla la media y la desviación típica.
- d) ¿Es la distribución aproximadamente normal?

10. Ejercicio 13 del tema 1

Los manatíes son unos animales grandes y dóciles que viven a lo largo de la costa de Florida. Cada año las lanchas motoras hieren o matan muchos de ellos. A continuación se presenta una tabla que contiene, para cada año, el número de licencias para motoras (expresado en miles de licencias) expedidas en Florida y el número de manatíes muertos en los años 1977 a 1990.

- a) Queremos analizar la relación entre el número de licencias expedidas anualmente en Florida y el número de manatíes muertos. ¿Cuál es la variable explicativa? Dibuja un diagrama de dispersión con esos datos. ¿Qué nos dice el diagrama sobre la relación entre esas dos variables? Las variables ¿están asociadas positiva o negativamente?
- b) Halla la recta de regresión para expresar el número de manatíes muertos en función del número de licencias.
- c) Describe la fuerza de la relación. ¿Se puede predecir con precisión el número de manatíes muertos cada año conociendo el número de licencias expedidas ese año? Si Florida decidiera congelar el número de licencias en 700.000, ¿cuántos manatíes matarían, aproximadamente, las lanchas motoras?

| Año | Licencias | Manatíes | Año | Licencias | Manatíes |
|------|-----------|----------|------|-----------|----------|
| 1977 | 447 | 13 | 1984 | 559 | 34 |
| 1978 | 460 | 21 | 1985 | 585 | 33 |
| 1979 | 481 | 24 | 1986 | 614 | 33 |
| 1980 | 498 | 16 | 1987 | 645 | 39 |
| 1981 | 513 | 24 | 1988 | 675 | 43 |
| 1982 | 512 | 20 | 1989 | 711 | 50 |
| 1983 | 526 | 15 | 1990 | 719 | 47 |