

## SPSS: DESCRIPTIVOS

### PROCEDIMIENTO DE ANÁLISIS INICIAL DE DATOS: FRECUENCIAS Y ESTADÍSTICOS

Cuando analizamos los datos, lo primero que tenemos que hacer es formar una idea sobre las características de las variables. Para eso hay que fijarse en los aspectos básicos de las variables, como tendencia central, dispersión y forma de distribución.

**Tiene mucha importancia el tipo de variable que estamos analizando,**

**Variable categórica:** las medidas de tendencia central y de dispersión carecen de importancia comparadas con la utilidad de una distribución de frecuencias o un gráfico de sobre la forma de distribución.

**Variable continua:** distribución de frecuencias pierde importancia comparada con la capacidad informativa de las medidas de tendencia central y de dispersión.

Los diagramas de distribución también tienen que ser diferentes dependiendo de que la variable estudiada sea categórica o continua.

### FRECUENCIAS:

Informa sobre los valores de la variable y el número de veces (porcentajes) que se repiten esos valores.

- Podemos activar opciones para poder calcular algunos de los estadísticos descriptivos más utilizados (tendencia central, posición, dispersión, asimetría, curtosis).
- Construir algunos gráficos (de barras, sectores, histogramas).
- Controlar el formato.

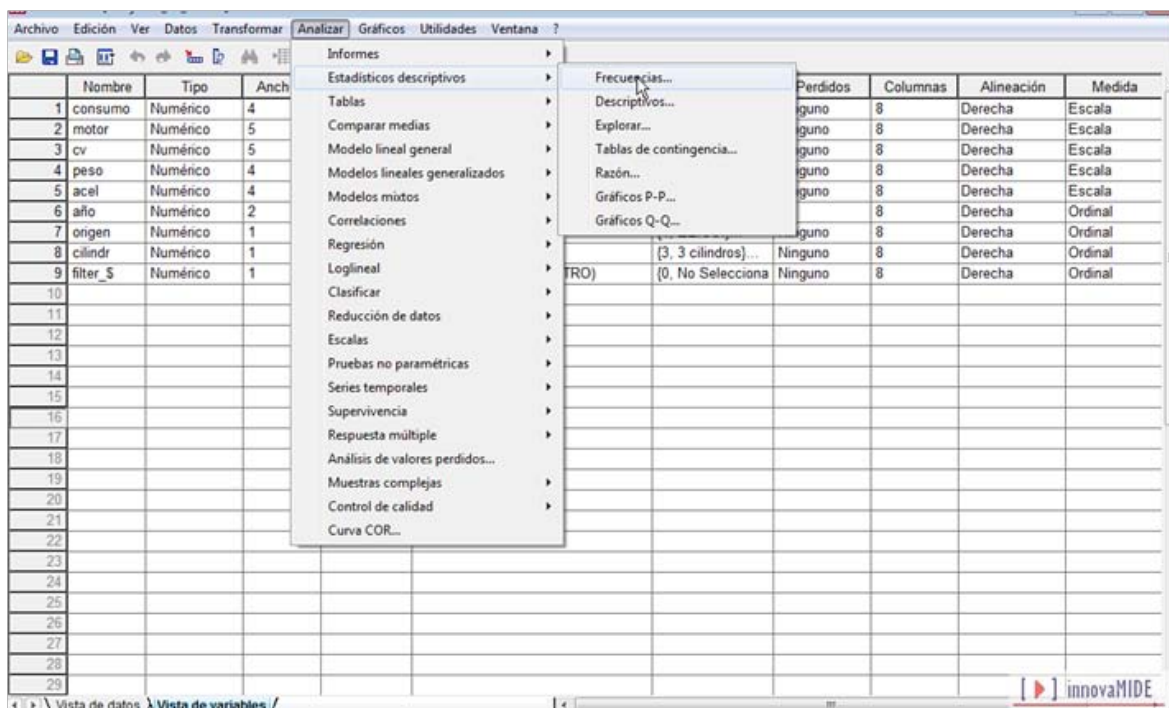
La utilización de las opciones depende del hecho de que la variable sea categórica o continua.

Para obtener distribución de frecuencias:

Menú: Analizar

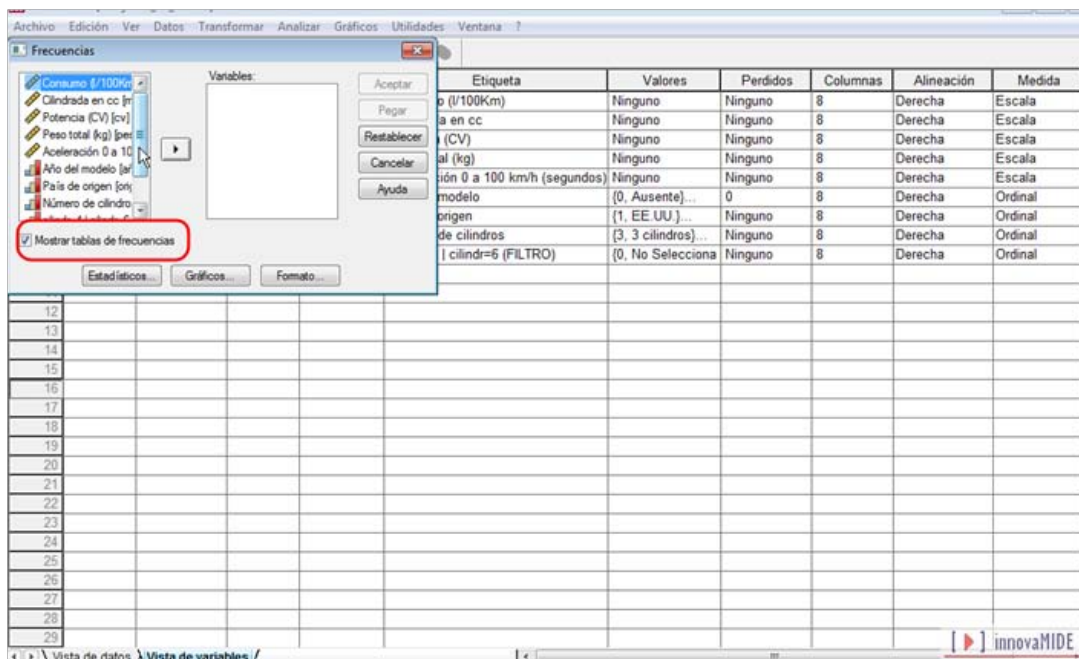
- Estadísticos Descriptivos

- Frecuencias...



En el primer cuadro que aparece seleccionamos las variables que nos interesa describir.

La opción **Mostrar tablas de frecuencias** está activada por defecto, y nos permite decidir si se desea o no obtener la distribución de frecuencias. Puede desactivarse si sólo queremos ver algún gráfico y algún estadístico descriptivo, los resultados mostrarán sólo el número total de casos y de valores perdidos.



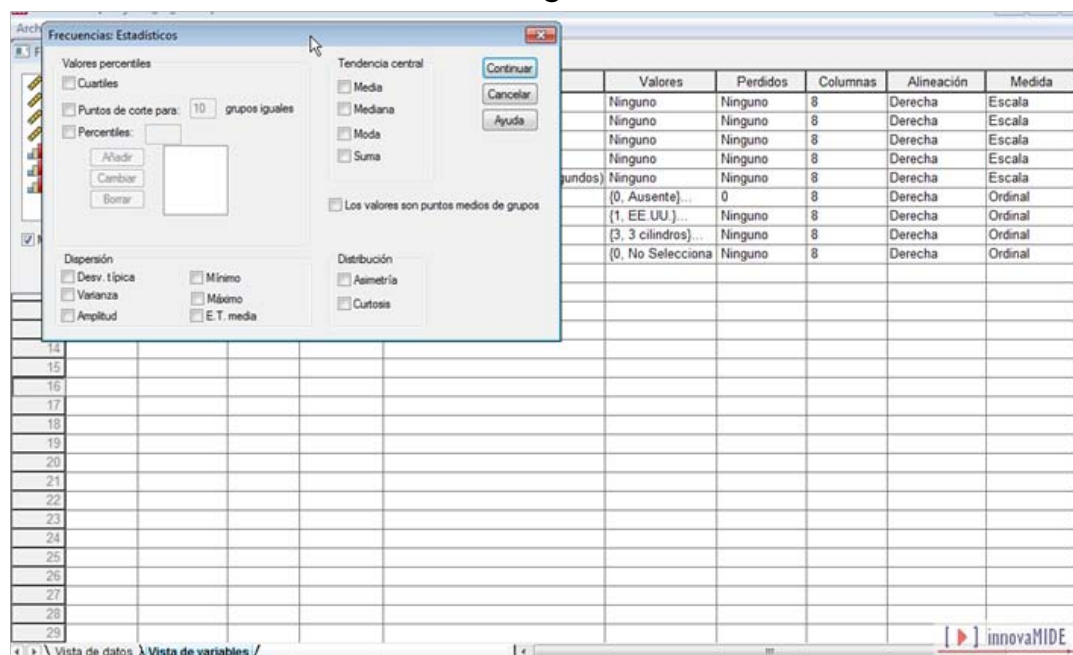
## ESTADISTICOS.

Podemos obtener información adicional utilizando los botones específicos del cuadro de diálogo Frecuencias. Para obtener estadísticos descriptivos:

Menú: Analizar

- Estadísticos...

En el subcuadro de diálogo Frecuencias: Estadísticos...



Ejemplo de tabla de Frecuencias para una variable categórica:

Archivo Edición Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ventana ?

FRECUENCIAS  
VARIABLES=cilindr  
/ORDER= ANALYSIS .

→ Frecuencias

[Conjunto\_de\_datos2] C:\Program Files (x86)\SPSS\Coche.sav

Estadísticos

Número de cilindros

N	Válidos	405
	Perdidos	1

Número de cilindros

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	3 cilindros	4	1,0	1,0	1,0
	4 cilindros	207	51,0	51,1	52,1
	5 cilindros	3	,7	,7	52,8
	6 cilindros	84	20,7	20,7	73,6
	8 cilindros	107	26,4	26,4	100,0
	Total	405	99,8	100,0	
Perdidos	Sistema	1	,2		
	Total	406	100,0		

SPSS El procesador está preparado

Estas tablas son las que se generan de forma automática (con la casilla Mostrar tablas de frecuencias activada) y nos muestra 2 tablas: Estadísticos, que incluye Número de casos válidos y perdidos y otra tabla que nos presenta datos de Frecuencias, Porcentajes, Porcentajes válidos y Porcentajes acumulados (en este caso para la variable “numero de cilindros”).

→ Frecuencias

[Conjunto\_de\_datos2] C:\Program Files (x86)\SPSS\Coche.sav

Estadísticos

Número de cilindros

N	Válidos	405
	Perdidos	1

Número de cilindros

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	3 cilindros	4	1,0	1,0	1,0
	4 cilindros	207	51,0	51,1	52,1
	5 cilindros	3	,7	,7	52,8
	6 cilindros	84	20,7	20,7	73,6
	8 cilindros	107	26,4	26,4	100,0
	Total	405	99,8	100,0	
Perdidos	Sistema	1	,2		
	Total	406	100,0		

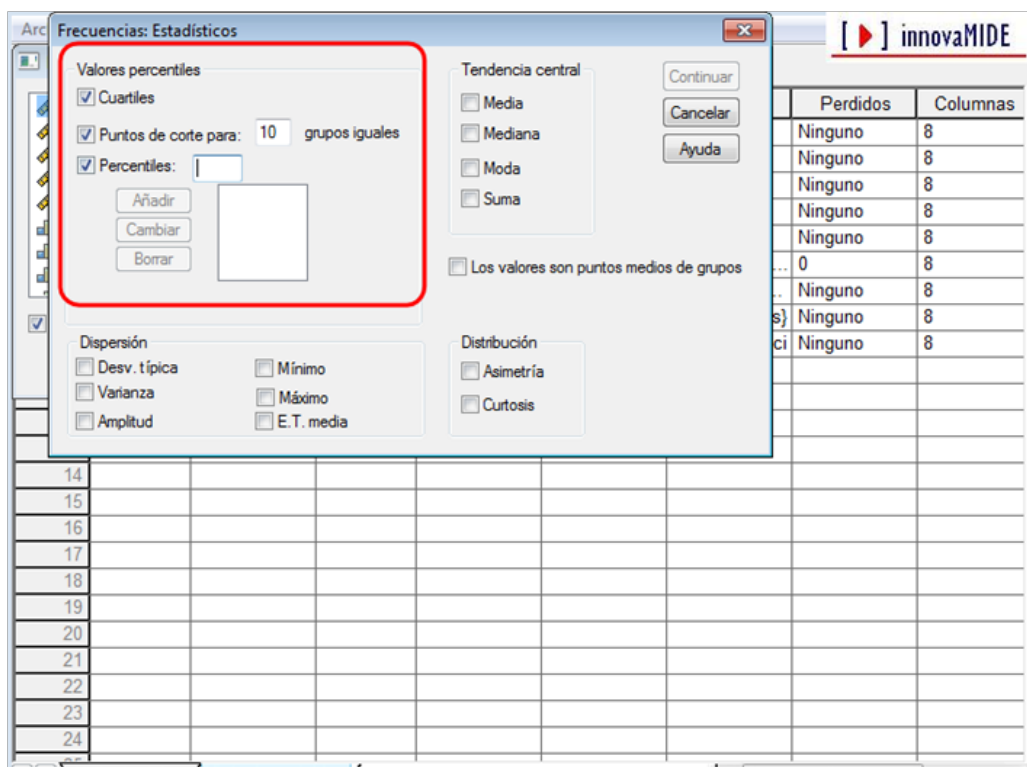
SPSS El procesador está preparado

En este caso la columna de la izquierda nos indica las categorías de la variable y las siguientes columnas presentan datos de Frecuencias, Porcentajes, Porcentajes válidos y Porcentajes acumulados para cada categoría de la variable.

- *Frecuencia* – es el número de veces que se repite un valor;
- *Porcentaje* – la frecuencia porcentual del valor en relación con otros valores de la variable;
- *Porcentaje válido* – la frecuencia porcentual, calculada sin tener en cuenta los casos perdidos;
- *Porcentaje acumulado* – la frecuencia porcentual, sumando el valor del porcentaje de la categoría anterior;
- La última línea muestra el número total de casos.

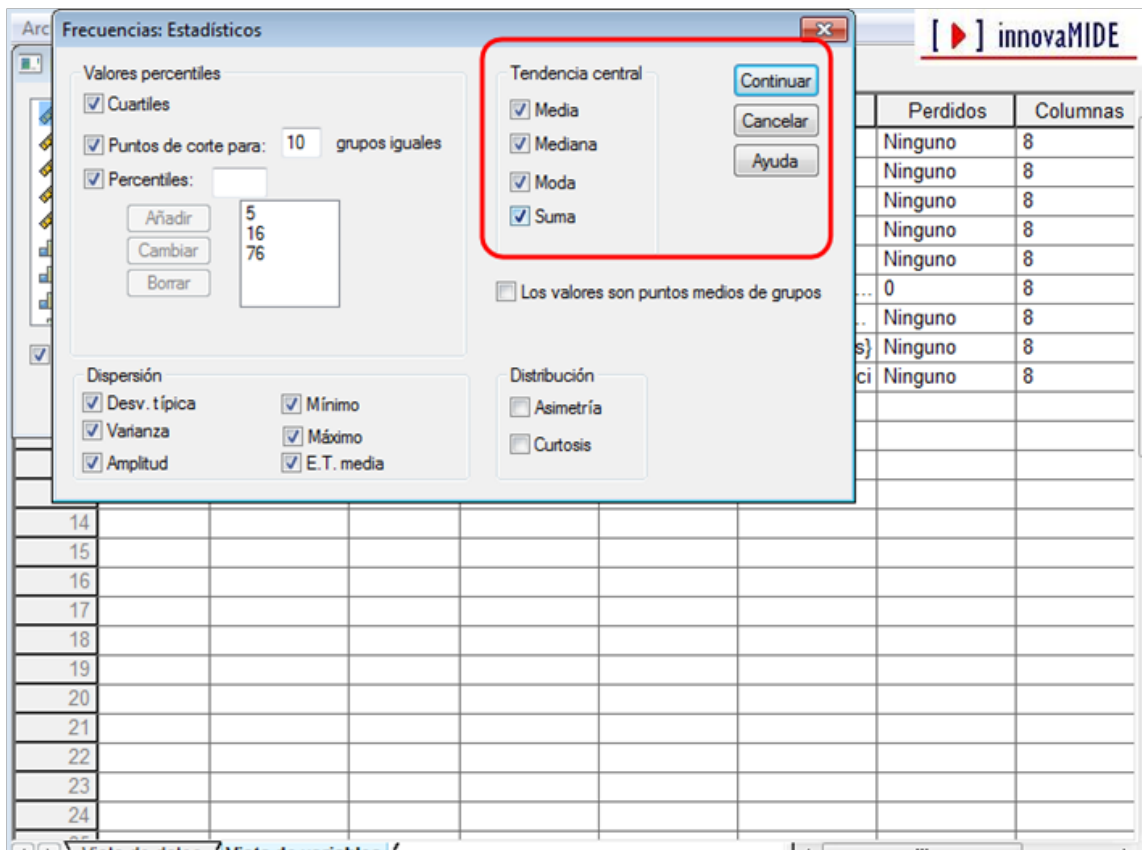
**De forma adicional podemos solicitar siguientes estadísticos:**

*Valores percentiles* (sólo tiene sentido si las variables son ordinales): podemos solicitar cuantiles: cuartiles, deciles, percentiles, etc.



- *Cuartiles*: percentiles 25, 50 y 75, es decir, los valores por debajo de los cuales se encuentra 25%, 50% y 75% de los casos respectivamente.
- *Puntos de corte para k grupos iguales*: calcula valores que dividen la muestra en grupos de mismo tamaño. El valor marcado por defecto es de 10, pero puede escribirse cualquier valor entre 2 y 100.
- *Percentiles*: podemos solicitar percentiles concretos (en nuestro ejemplo hemos solicitado percentiles 5, 16 y 76).

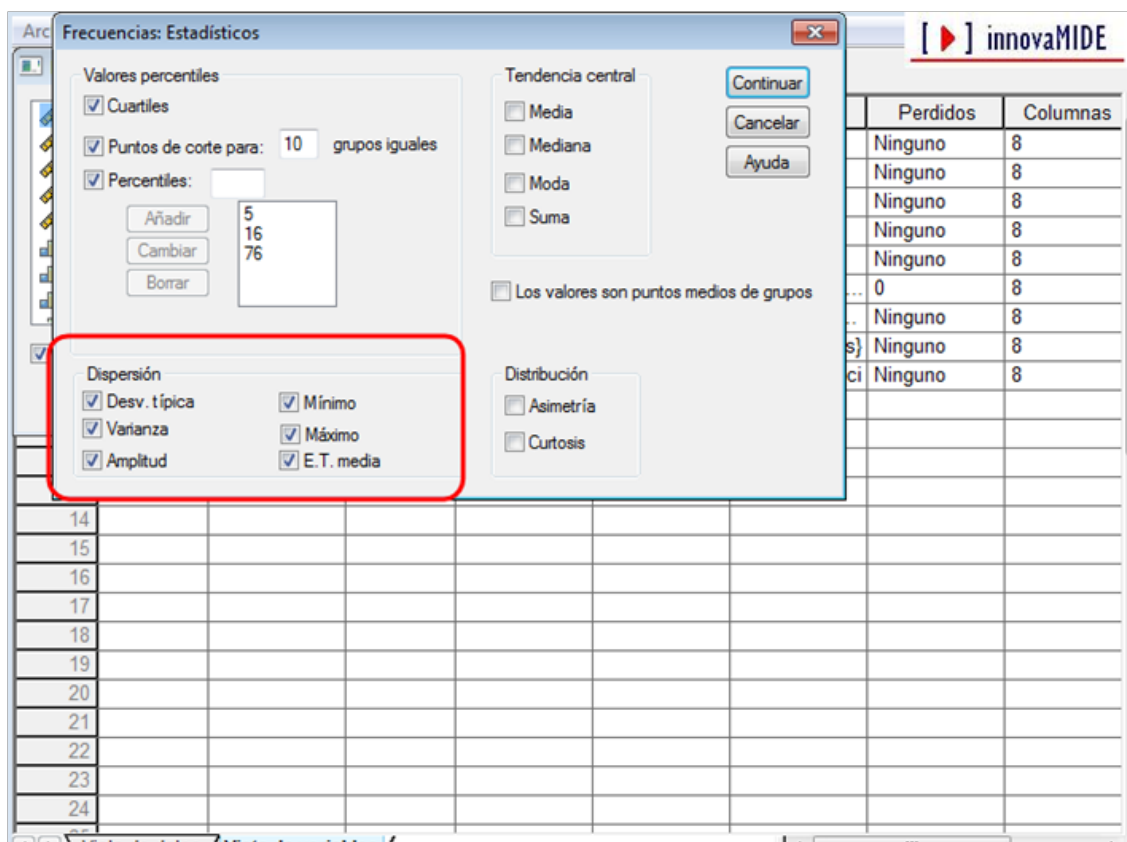
Tendencia central: podemos seleccionar uno o más de los siguientes estadísticos:



- *Media*: media aritmética. Requiere variables cuantitativas (de intervalo o razón, aunque también podemos calcularla con datos ordinales).

- *Mediana*: percentil 50. Si el número de casos es par, la mediana se calcula como el promedio de los dos casos centrales cuando éstos se encuentran ordenados. Si el número de casos es impar, la mediana es el valor del caso central de la distribución. La mediana es un estadístico típicamente ordinal (requiere variables ordinales al menos).
- *Moda*. El valor que más se repite. Si existen 2 o más valores empatados en el número de repeticiones, solo se muestra el más pequeño de ellos. Sirve para todo tipo de variables, pero es más apropiada para datos categóricos (es el estadístico que sólo aprovecha la información nominal y con las variables continuas es esperable que todos los valores tengan una frecuencia igual a 1).
- *Suma*. Suma de todos los valores.

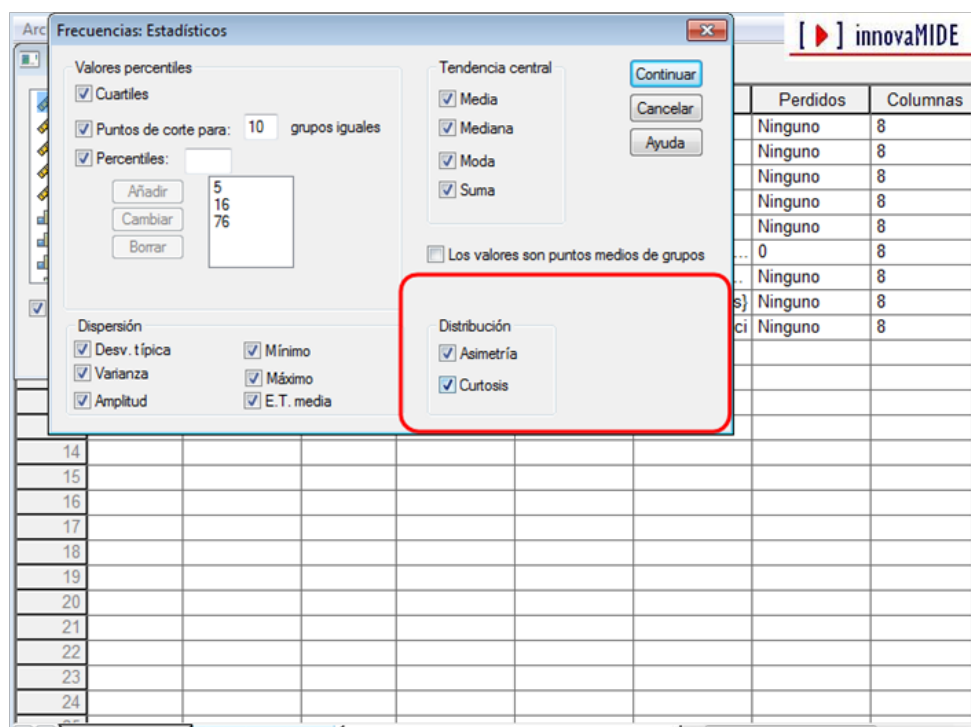
Dispersión: podemos seleccionar uno o más de los estadísticos:





- *Desviación típica* (apropiado para variables cuantitativas): raíz cuadrada de la varianza. Mide el grado en que las puntuaciones de la variable se alejan de su media.
- *Varianza* (apropiado para variables cuantitativas): medida de dispersión que se obtiene dividiendo por  $n-1$  la suma de los cuadrados de las diferencias entre cada puntuación y la media.
- *Amplitud* (no apropiado para las variables nominales): diferencia entre el valor máximo y mínimo.
- *Mínimo*.
- *Máximo*.
- *E.T. media* (apropiado para variables cuantitativas): Error típico de la media: desviación típica de la distribución muestral de la media. Se obtiene dividiendo la desviación típica de la variable por la raíz cuadrada del número de casos.

Distribución. Podemos seleccionar uno o más estadísticos.



- *Asimetría* (sólo tiene sentido para variables cuantitativas): índice que expresa el grado de asimetría de la distribución. La asimetría positiva indica que los valores más extremos se encuentran por encima de la media. La asimetría negativa



indica que los valores más extremos se encuentran por debajo de la media. Los índices de asimetría próximos a cero indican simetría. Los resultados también recogen el error típico del índice de asimetría - la desviación típica de la distribución muestral del índice de asimetría, el cual permite tipificar el valor del índice de asimetría e interpretarlo como una puntuación  $z$  con distribución aproximadamente  $N(0,1)$ . Índices tipificados mayores que 1,96 en valor absoluto permiten afirmar que existe asimetría (positiva o negativa, dependiendo del signo del índice).

- **Curtosis** (sólo tiene sentido para variables cuantitativas): índice que expresa el grado en que una distribución acumula casos en sus colas en comparación con los casos acumulados en las colas de una distribución normal con la misma varianza. La curtosis positiva indica que en las colas de la distribución hay acumulados más casos que en las colas de una distribución normal (lo cual suele coincidir con distribuciones más puntiagudas que una distribución normal). Los índices de curtosis próximos a cero indican semejanza con la curva normal.

Los resultados también recogen el error típico del índice de curtosis, el cual puede utilizarse para tipificar el valor del índice de curtosis y poder interpretarlo como una puntuación  $z$  distribuida aproximadamente  $N(0,1)$ . Índices mayores que 1,96 en valor absoluto permiten afirmar que la distribución se aleja de la distribución normal.

Los valores son puntos medios de grupos: si la variable que deseamos estudiar se encuentra agrupada en intervalos, esta opción permite calcular la mediana y los percentiles, considerando que los valores de la variable son los puntos medios de intervalos uniformemente distribuidos. Esta opción afecta a todas las variables de la lista Variables, no debería marcarse si una o más variables de las listadas no se encuentran agrupadas en intervalos.

*El material de este manual fue adaptado a partir de la información del Tutorial de SPSS15.*