

Capítulo 10

Análisis descriptivo: Los procedimientos *Frecuencias* y *Descriptivos*

Al analizar datos, lo primero que conviene hacer con una variable es, generalmente, formarse una idea lo más exacta posible acerca de sus características. Esto se consigue prestando atención a tres aspectos básicos: tendencia central, dispersión y forma de la distribución. Ahora bien, las medidas de tendencia central y de dispersión, y los índices y gráficos sobre la forma de la distribución, resultan más o menos útiles dependiendo del tipo de variable que se intente caracterizar. Con variables categóricas, por ejemplo, las medidas de tendencia central y de dispersión carecen de importancia comparadas con la utilidad de una distribución de frecuencias o un gráfico sobre la forma de la distribución. Por el contrario, con variables continuas una distribución de frecuencias pierde importancia comparada con la capacidad informativa de las medidas de tendencia central y de dispersión. Por otro lado, los diagramas que informan sobre la forma de una distribución son diferentes dependiendo de que la variable estudiada sea categórica o continua.

En este capítulo vamos a describir dos procedimientos SPSS que permiten obtener la información necesaria para caracterizar apropiadamente tanto variables categóricas como cuantitativas: el procedimiento **Frecuencias** y el procedimiento **Descriptivos**.

Frecuencias

Una distribución de frecuencias informa sobre los valores concretos que adopta una variable y sobre el número (y porcentaje) de veces que se repite cada uno de esos valores. El procedimiento **Frecuencias** permite obtener distribuciones de frecuencias, pero además contiene opciones para:

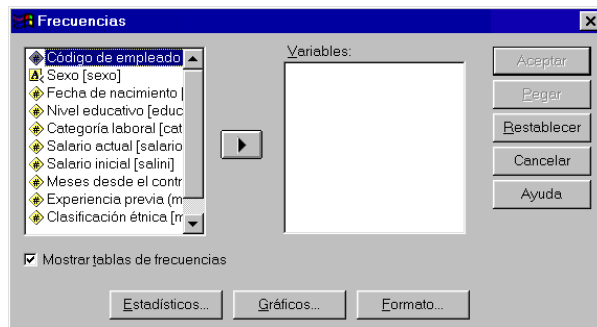
- Calcular algunos de los estadísticos descriptivos más utilizados (sobre tendencia central, posición, dispersión, asimetría y curtosis).
- Construir algunos diagramas básicos (gráficos de barras, de sectores e histogramas).
- Controlar el formato de presentación de las distribuciones de frecuencias.

La utilización de estas opciones depende en gran medida del hecho de que la variable estudiada sea categórica o continua.

Para obtener una distribución de frecuencias:

- ▶ Seleccionar la opción **Estadísticos descriptivos > Frecuencias...** del menú **Analizar** para acceder al cuadro de diálogo *Frecuencias* que muestra la figura 10.1.

Figura 10.1. Cuadro de diálogo *Frecuencias*.



Este cuadro de diálogo permite obtener distribuciones de frecuencias absolutas y porcentuales, varios estadísticos descriptivos y algunos gráficos básicos. Para ello:

- ▶ Trasladar a la lista **Variables**, mediante el botón flecha, la variable o variables cuya distribución de frecuencias se desea obtener. La especificación mínima requerida es una variable numérica o de cadena corta; las variables de cadena larga no están disponibles en la lista de variables del archivo de datos.
- ▶ Pulsar el botón **Aceptar**.
- Mostrar tablas de frecuencias.** Esta opción (activa por defecto) permite decidir si se desea o no obtener la distribución de frecuencias. Puede desactivarse si, por ejemplo, sólo interesa ver algún gráfico o algún estadístico descriptivo (como suele ocurrir cuando se intenta describir una variable cuantitativa continua).

Si se desactiva esta opción y no se efectúa ninguna otra selección, los resultados sólo muestran el número total de casos y de valores perdidos.

Ejemplo (Estadísticos descriptivos > Frecuencias)

Este ejemplo muestra cómo obtener una distribución de frecuencias con las especificaciones mínimas que el procedimiento **Frecuencias** tiene establecidas por defecto:

- ▶ Seleccionar la opción **Estadísticos descriptivos > Frecuencias** del menú **Analizar**.
- ▶ Trasladar, mediante el botón *flecha*, la variable *catlab* (categoría laboral) a la lista **Variables**.
- ▶ Pulsar el botón **Aceptar**.

Aceptando estas elecciones, el *Visor de resultados* ofrece dos tablas pivotantes: una con el número de casos válidos y de casos con valores perdidos (tabla 10.1.a), y otra con la distribución de frecuencias solicitada (tabla 10.1.b).

Figura 10.1.a. Tabla de estadísticos del procedimiento *Frecuencias*.

	N	
	Válidos	Perdidos
Categoría laboral	474	0

Tabla 10.1.b. Tabla de frecuencias del procedimiento *Frecuencias*.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Administrativo	363	76,6	76,6	76,6
	Seguridad	27	5,7	5,7	82,3
	Directivo	84	17,7	17,7	100,0
	Total	474	100,0	100,0	
Total		474	100,0		

Puesto que no hemos seleccionado ningún estadístico, la tabla 10.1.a sólo muestra el número de casos válidos y el número de valores perdidos.

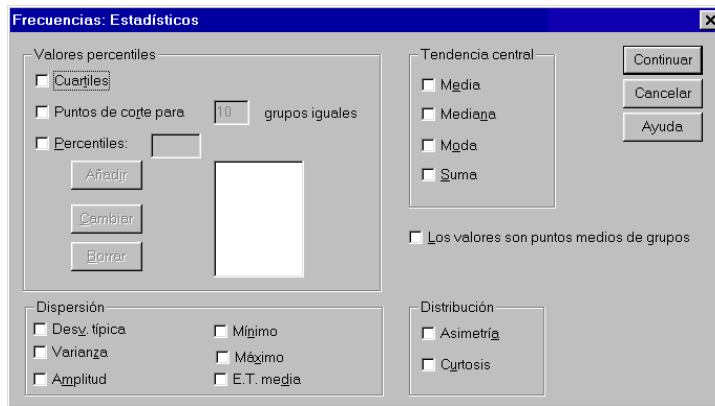
La tabla 10.1 ofrece información sobre: los valores de la variable *catlab* (sus etiquetas), la frecuencia absoluta de cada valor (*Frecuencia*), la frecuencia porcentual (*Porcentaje*), la frecuencia porcentual calculada sobre los casos válidos, es decir, sin tener en cuenta los casos con valores perdidos (*Porcentaje válido*) y la frecuencia porcentual acumulada (*Porcentaje acumulado*). En la última línea aparece el número total de casos.

Estadísticos

Podemos obtener información adicional utilizando los botones específicos del cuadro de diálogo *Frecuencias* (figura 10.1). Para obtener estadísticos descriptivos:

- ▶ Pulsar el botón **Estadísticos...** para acceder al subcuadro de diálogo *Frecuencias: Estadísticos* que muestra la figura 10.2.

Figura 10.2. Subcuadro de diálogo *Frecuencias: Estadísticos*.



Valores percentiles. Este recuadro contiene varias opciones para solicitar cuantiles: cuartiles, deciles, percentiles, etc.

- Cuartiles.** Calcula los percentiles 25, 50 y 75, es decir, los valores por debajo de los cuales se encuentra el 25 %, el 50 % y el 75 % de los casos, respectivamente. Para saber cómo calcula el SPSS estos cuantiles, puede consultarse el apartado *Estadísticos* del capítulo 11.
- Puntos de corte para k grupos iguales.** Calcula los $k-1$ valores que dividen la muestra en k grupos del mismo tamaño. El valor por defecto de k es 10, pero puede escribirse cualquier otro valor entre 2 y 100.

- Percentiles.** Permite solicitar percentiles concretos (valores que acumulan un determinado porcentaje de casos). Para obtener un percentil concreto:
 - Escribir el valor deseado en el cuadro de texto **Percentiles**.
 - Pulsar el botón **Añadir** para trasladar ese valor a la lista de percentiles.
 - Pulsar los botones **Cambiar** y **Borrar** para modificar o eliminar, respectivamente, valores previamente añadidos.

Tendencia central. Puede seleccionarse uno o más de los siguientes estadísticos:

- Media.** Media aritmética: suma de todas las puntuaciones dividida por el número de puntuaciones.
- Mediana.** Valor por debajo del cual se encuentra el 50 % de los casos (equivale al 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.
- Moda.** Valor que más se repite. Si existen dos o más valores empatados en el número de repeticiones, solo se muestra el más pequeño de ellos.
- Suma.** Suma de todos los valores.

Dispersión. Puede seleccionarse uno o más de los siguientes estadísticos:

- Desv. típica.** Desviación típica: raíz cuadrada de la varianza. Mide el grado en que las puntuaciones de la variable se alejan de su media.
- Varianza.** 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.** Diferencia entre el valor más grande (máximo) y el más pequeño (mínimo).
- Mínimo.** Valor más pequeño.
- Máximo.** Valor más grande.
- E.T. media.** 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. Puede seleccionarse uno o más de los siguientes estadísticos:

- Asimetría.** 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 (es decir, 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.** 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.** En el caso de que la variable que deseamos estudiar se encuentre agrupada en intervalos, esta opción permite calcular la mediana y los percentiles interpolando valores, es decir, considerando que los valores de la variable son los puntos medios de intervalos uniformemente distribuidos.

Puesto que esta opción afecta a todas las variables de la lista **Variables** (ver figura 10.1), no debería marcarse si una o más variables de las listadas no se encuentran agrupadas en intervalos.

Cuándo utilizar cada estadístico

Por lo que se refiere a los *percentiles*, sólo tiene sentido calcularlos con variables al menos ordinales. Carecen de significado con variables nominales.

Entre las medidas de *tendencia central*, la media requiere variables cuantitativas (de intervalo o razón, aunque también suele calcularse con datos ordinales). La mediana es un estadístico típicamente ordinal (requiere variables ordinales al menos). Al contrario de lo que ocurre con la media, la mediana es insensible a la presencia de valores extremos y, por tanto, es preferible a la media cuando la distribución es asimétrica. La moda sirve para todo tipo de variables, pero es más apropiada para caracterizar datos categóricos porque, por un lado, es un estadístico que sólo aprovecha información nominal y, por otro, con variables continuas es esperable que todos los valores tengan una frecuencia igual a 1.

En cuanto a las medidas de *dispersión*, la desviación típica, la varianza y el error típico de la media únicamente poseen significado con variables cuantitativas (de intervalo o razón, aunque también suelen calcularse con datos ordinales). La amplitud o rango es apropiada para todo tipo de variables, excepto para las nominales, en las que no tiene sentido hablar de dispersión.

En lo relativo a los índices de *asimetría y curtosis*, de nuevo solo tiene sentido calcularlos con variables cuantitativas.

Ejemplo (Estadísticos descriptivos > Frecuencias > Estadísticos)

Este ejemplo muestra cómo obtener algunos estadísticos descriptivos utilizando el procedimiento **Frecuencias**:

- ▶ En el cuadro de diálogo *Frecuencias* (ver figura 10.1), seleccionar la variable *salario* y trasladarla a la lista **Variables**.
- ▶ Pulsar el botón **Estadísticos...** y marcar todas las opciones del subcuadro de diálogo *Frecuencias: Estadísticos* (ver figura 10.2).

Con estas elecciones, el *Visor* muestra los estadísticos que recoge la tabla 10.2 (la tabla ha sido *pivotada* para adaptarla mejor a las dimensiones de la página):

Tabla 10.2. Tabla de *Estadísticos* del procedimiento *Frecuencias*.

Salario actual

		Estadístico	Error típ.
N	Válidos	474	
	Perdidos	0	
Media		\$34,419.57	\$784.31
Mediana		\$28,875.00	
Moda		\$30,750	
Desv. típ.		\$17,075.66	
Varianza		\$291,578,214.45	
Asimetría		2,125	,112
Curtosis		5,378	,224
Rango		\$119,250	
Mínimo		\$15,750	
Máximo		\$135,000	
Suma		\$16,314,875	
Percentiles	\$10.00	\$21,000.00	
	\$20.00	\$22,950.00	
	\$25.00	\$24,000.00	
	\$30.00	\$24,825.00	
	\$40.00	\$26,700.00	
	\$50.00	\$28,875.00	
	\$60.00	\$30,750.00	
	\$70.00	\$34,500.00	
	\$75.00	\$37,162.50	
	\$80.00	\$41,100.00	
\$90.00	\$59,700.00		

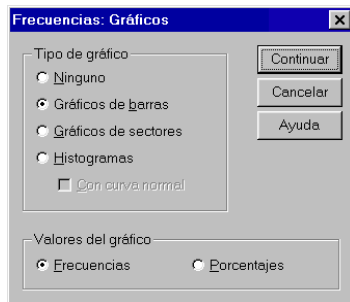
Observando la tabla vemos, por ejemplo, que el salario medio es de 34.419,57 dólares (*Media*), que la mitad de los sujetos tienen salarios por debajo de 28.875 dólares (*Mediana, Percentil 50*), que entre el sujeto que gana más y el que gana menos existe una diferencia de 119.250 dólares (*Rango*), que el 50 % de los sujetos tiene salarios comprendidos entre 24.000 y 37.162,50 dólares (*Percentiles 25 y 75*), que el grado de asimetría es acusadamente positivo (pues el cociente entre el índice de asimetría y su error típico vale $2,125/0,112 = 18,97$, y este valor es demasiado grande para pensar que pertenece a una distribución con valor esperado cero), y que la acumulación de casos en las colas (*Curtosis*) es mayor que la que corresponde a una distribución normal (pues tipificando el índice de curtosis obtenemos $5,378/0,224 = 24,01$, y este valor es demasiado grande para pensar que pertenece a una distribución con valor esperado cero, que es el valor que indica una curtosis equivalente a la de una curva *normal*).

Gráficos

El procedimiento **Frecuencias** también ofrece la posibilidad de obtener algunos gráficos básicos. En concreto, permite obtener gráficos de *barras*, gráficos de *sectores* e *histogramas*. (El menú **Gráficos** de la barra de menús también permite obtener estos mismos gráficos, además de otros muchos). Para obtener un gráfico de *barras*, un gráfico de *sectores*, o un *histograma*:

- ▶ Pulsar el botón **Gráficos...** del cuadro de diálogo *Frecuencias* (ver figura 10.1) para acceder al subcuadro de diálogo *Frecuencias: Gráficos* que muestra la figura 10.3.

Figura 10.3. Subcuadro de diálogo *Frecuencias: Gráficos*.



Tipo de gráfico. Puede elegirse entre:

- **Ninguno.** No se genera ningún gráfico. Es la opción por defecto.
- **Gráficos de barras.** Gráficos en los que a cada valor de la variable se le asigna una barra con altura equivalente a su frecuencia absoluta o porcentual. La escala de la altura de las barras se ajusta automáticamente teniendo en cuenta la frecuencia más alta de las representadas.
- **Gráficos de sectores.** Gráficos circulares en los que a cada valor de la variable se le asigna un sector de tamaño equivalente a su frecuencia absoluta o porcentual.
- **Histogramas.** Similares a los gráficos de barras, pero con las barras juntas, dando así la impresión de continuidad. Sólo pueden obtenerse con variables numéricas. Para construir el histograma, el SPSS agrupa la variable en 21 intervalos (o menos, si la amplitud de la variable es menor que 21).

- Con curva normal.** Esta opción permite obtener una curva normal superpuesta sobre el histograma (la curva normal se genera a partir de la media y la desviación típica de la variable representada).

Valores del gráfico. En los gráficos de barras y de sectores es posible decidir qué tipo de frecuencia se desea representar:

- Frecuencias.** La escala y la etiqueta del eje correspondiente a la altura de las barras (o al tamaño de los sectores) están expresadas en frecuencias absolutas. Es la opción por defecto.
- Porcentajes.** La escala y la etiqueta del eje correspondiente a la altura de las barras (o al tamaño de los sectores) están expresadas en frecuencias porcentuales.

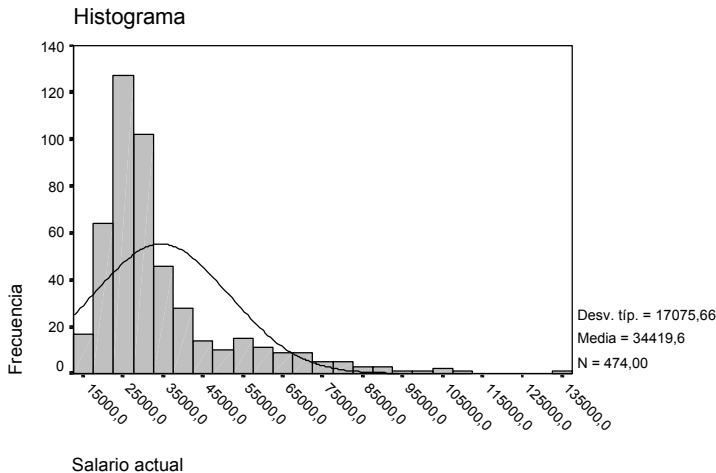
Ejemplo (Estadísticos descriptivos > Frecuencias > Gráficos)

Este ejemplo muestra cómo obtener un histograma de la variable *salario* utilizando la opción **Gráficos...** del procedimiento **Frecuencias**.

- ▣ En el cuadro de diálogo *Frecuencias* (ver figura 10.1), trasladar la variable *salario* a la lista **Variables**.
- ▣ Pulsar el botón **Gráficos...**
- ▣ En el subcuadro de diálogo *Frecuencias: Gráficos* (ver figura 10.3) seleccionar **Histogramas** y marcar la opción **Con curva normal**.

Aceptando estas elecciones, obtenemos el histograma que muestra la figura 10.4.

Figura 10.4. Histograma de la variable *salario*.



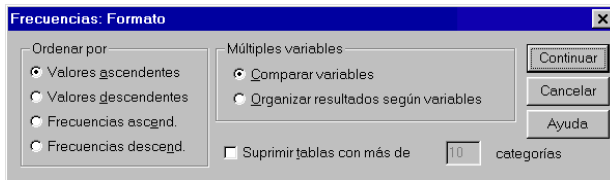
Tal como habíamos anticipado ya basándonos en los índices de asimetría y curtosis del ejemplo anterior, el histograma muestra asimetría positiva, al tiempo que una evidente desviación de la normalidad.

Formato

Las opciones de formato permiten controlar algunos aspectos relacionados con la forma en que aparecerán en el *Visor* las tablas de frecuencias y los estadísticos solicitados. Para controlar el formato de presentación de las tablas de frecuencias:

- ▶ Pulsar el botón **Formato...** del cuadro de diálogo *Frecuencias* (ver figura 10.1) para acceder al subcuadro de diálogo *Frecuencias: Formato* que muestra la figura 10.5.

Figura 10.5. Cuadro de diálogo *Frecuencias: Formato*.



Ordenar por. Las opciones de este recuadro sirven para establecer el orden en el que aparecerán los valores o categorías de la variable en la distribución de frecuencias:

- **Valores ascendentes.** Los valores o categorías de la variable son ordenados desde el más pequeño al más grande. Es la opción por defecto.
- **Valores descendentes.** Los valores o categorías de la variable son ordenados desde el más grande al más pequeño.
- **Frecuencias ascendentes.** Los valores o categorías de la variable se ordenan de forma ascendente tomando como criterio el tamaño de cada frecuencia.
- **Frecuencias descendentes.** Los valores o categorías de la variable se ordenan de forma descendente tomando como criterio el tamaño de cada frecuencia.

Si se solicita algún percentil o algún histograma, los valores se ordenan de forma ascendente (la opción por defecto) independientemente del orden seleccionado.

- **Suprimir tablas con más de k categorías.** Esta opción elimina de la salida las distribuciones con más de k valores o categorías. El valor por defecto para k es 10, pero puede introducirse cualquier valor igual o mayor que 1.

Esta opción es particularmente útil si en la lista **Variables** (figura 10.1) se han seleccionado variables continuas y categóricas: permite eliminar selectivamente las distribuciones de frecuencias de las variables continuas al tiempo que permite obtener las de las variables categóricas. (Generalmente, con variables cuantitativas continuas interesa obtener estadísticos descriptivos o histogramas, pero no distribuciones de frecuencias, pues éstas son demasiado largas y, por tanto, poco informativas).

Múltiples variables. Al solicitar gráficos o estadísticos para más de una variable:

- Comparar variables** utiliza una sola tabla de resultados para todas las variables.
- Organizar resultados según variables** utiliza una tabla para cada variable.

Descriptivos

A diferencia de lo que ocurre con el procedimiento **Frecuencias**, que contiene opciones para describir tanto variables categóricas como cuantitativas continuas, el procedimiento **Descriptivos** está diseñado únicamente para variables cuantitativas continuas. Contiene unos cuantos estadísticos descriptivos (tendencia central, dispersión y forma de la distribución) que también incluye el procedimiento **Frecuencias**, pero añade una opción especialmente importante: la posibilidad de obtener *puntuaciones típicas*. Para obtener estadísticos descriptivos y puntuaciones típicas:

- ▶ Seleccionar la opción **Estadísticos descriptivos > Descriptivos** del menú **Analizar** para acceder al cuadro de diálogo *Descriptivos* que muestra la figura 10.6.

Figura 10.6. Cuadro de diálogo *Descriptivos*.



Para obtener los estadísticos básicos (media aritmética, desviación típica, valor mínimo y valor máximo) que el procedimiento **Descriptivos** ofrece por defecto:

- ▶ Trasladar una o más variables a la lista **Variables**. La especificación mínima requerida es una variable numérica; las variables de cadena no están disponibles en la lista de variables del archivo de datos.
- ▶ Pulsar el botón **Aceptar**.

- **Guardar valores tipificados como variables.** Los valores tipificados, también llamados puntuaciones típicas o puntuaciones z (z scores), expresan el número de desviaciones típicas que cada valor se aleja de su media (ver, dentro de este mismo apartado *Descriptivos*, el sub-apartado *Puntuaciones típicas y curva normal*). Marcando esta opción, el SPSS crea en el archivo de datos del *Editor de datos* una nueva variable con las puntuaciones típicas correspondientes a cada caso. Esta nueva variable recibe, por defecto, el nombre de la variable original con el prefijo z (si el nombre de la variable original posee 8 caracteres, el nuevo nombre sólo utiliza los 7 primeros). Si, por ejemplo, solicitamos las puntuaciones típicas de la variable *salario*, el SPSS recoge esas puntuaciones típicas en una nueva variable a la que asigna el nombre *zsalario*. Si la variable *zsalario* ya existe, entonces el nombre asignado a la nueva variable es *zsc001* (y así sucesivamente hasta *zsc099*). El prefijo *zsc* proviene de z score.

La etiqueta de la nueva variable se obtiene a partir de la etiqueta de la variable original. Si la variable original posee etiqueta, la etiqueta de la nueva variable se forma con los primeros 31 caracteres de la etiqueta original acompañados del prefijo *zscore*.

Si el nombre de la nueva variable no contiene parte del nombre original (lo que ocurre cuando el nombre asignado es, por ejemplo, *z001*), la etiqueta de la nueva variable se forma con *zscore(nombre-de-la-variable-original)* y los primeros 31 caracteres de la etiqueta original.

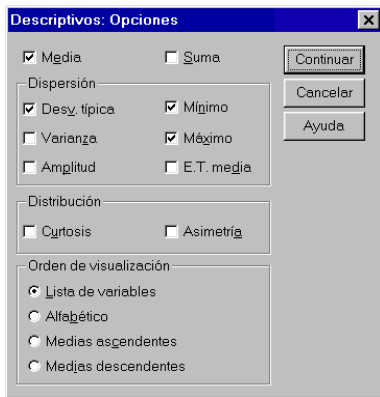
Si la variable original no posee etiqueta, la etiqueta de la nueva variable se forma con el nombre de la variable original acompañado del prefijo *zscore*.

Opciones

Para decidir qué estadísticos descriptivos se desea obtener y para controlar algunos aspectos de la presentación:

- ▶ Pulsar el botón **Opciones...** del cuadro de diálogo *Descriptivos* (ver figura 10.6) para acceder al subcuadro de diálogo *Descriptivos: Opciones* que muestra la figura 10.7.

Figura 10.7. Subcuadro de diálogo *Descriptivos: Opciones*.



- Media.** Media aritmética: suma de todas las puntuaciones dividida por el número de puntuaciones.
- Suma.** Suma de todos los valores.

Dispersión. Puede seleccionarse uno o más de los siguientes estadísticos de dispersión:

- Desv. típica.** Desviación típica: raíz cuadrada de la varianza. Mide el grado en que las puntuaciones de una variable se alejan de su media.
- Varianza.** 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.** Diferencia entre el valor más grande (máximo) y el más pequeño (mínimo).
- Mínimo.** Valor más pequeño.
- Máximo.** Valor más grande.
- E.T. media.** 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. Puede seleccionarse uno o más de los siguientes estadísticos:

- Curtosis.** 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 un mayor *apuntamiento*). Los índices de asimetría próximos a cero indican apuntamiento similar al de 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 e 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.

- Asimetría.** Índice que expresa el grado de asimetría de la distribución. La asimetría positiva indica que los valores más extremos tienden a situarse por encima de la media. La asimetría negativa indica que los valores más extremos tienden a situarse por debajo de la media. Los valores en torno a cero indican simetría.

Los resultados también recogen el *error típico* del índice de asimetría, el cual puede utilizarse para tipificar el valor del índice de asimetría e interpretarlo como una puntuación z distribuida 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).

Orden de visualización. Esta opción permite establecer el orden en el que serán listadas las variables en la tabla de descriptivos que ofrece el *Visor*:

- Lista de variables.** Las variables aparecen listadas en el mismo orden en el que han sido seleccionadas en el cuadro de diálogo *Descriptivos*, es decir, en el mismo orden que aparecen en el listado **Variables** de la figura 10.6. Es la opción por defecto.
- Alfabético.** Las variables aparecen listadas en orden alfabético.
- Medias ascendentes.** Las variables se ordenan por el tamaño de sus medias, desde la más pequeña hasta la más grande.
- Medias descendentes.** Las variables se ordenan por el tamaño de sus medias, desde la más grande hasta la más pequeña.

Ejemplo (Estadísticos descriptivos > Descriptivos > Opciones)

Este ejemplo muestra cómo obtener puntuaciones típicas y algunos estadísticos descriptivos utilizando el procedimiento **Descriptivos**:

- ▶ En el cuadro de diálogo *Descriptivos* (figura 10.6), trasladar las variables *salini*, *salario* y *tiempemp* a la lista **Variables**.
- ▶ Marcar la opción **Guardar valores tipificados como variables**.
- ▶ Pulsar el botón **Opciones...**
- ▶ En el subcuadro de diálogo *Descriptivos: Opciones* (figura 10.7) seleccionar las opciones **Media** y **Suma**, todas las opciones de los apartados **Dispersión** y **Distribución** y la opción **Lista de variables** del apartado **Orden de visualización**.

Aceptando estas elecciones, el *Visor* ofrece la información que muestra la tabla 10.3 (hemos pivotado la tabla para poder ajustarla mejor al tamaño de la página).

Comparando estos resultados con los obtenidos en el ejemplo referido al procedimiento **Frecuencias** (ver tabla 10.2), podemos observar que la diferencia radica en que, además de los estadísticos que ofrece el procedimiento **Descriptivos** (todos ellos apropiados para describir variables cuantitativas continuas), el procedimiento **Frecuencias** ofrece otros estadísticos susceptibles de ser utilizados con variables nominales y ordinales (moda, mediana, percentiles, etc).

Puede comprobarse en el *Editor de datos* que, además de los estadísticos descriptivos de la tabla 10.3, el procedimiento **Descriptivos** ha generado tres nuevas variables (con los nombres *zsalini*, *zsalario* y *ztiempem*) en el archivo *Datos de empleados*; estas tres nuevas variables contienen las puntuaciones típicas de las tres variables seleccionadas en el ejemplo: *salini*, *salario* y *tiempemp*.

Tabla 10.3. Tabla de *Estadísticos descriptivos* del procedimiento *Descriptivos*.

		Estadístico	Error típico
Salario inicial	N	474	
	Rango	\$70,980	
	Mínimo	\$9,000	
	Máximo	\$79,980	
	Suma	\$8,065,625	
	Media	\$17,016.09	\$361.51
	Desv. típ.	\$7,870.64	
	Varianza	61946944,959	
	Asimetría	2,853	,112
	Curtosis	12,390	,224
Salario actual	N	474	
	Rango	\$119,250	
	Mínimo	\$15,750	
	Máximo	\$135,000	
	Suma	\$16,314,875	
	Media	\$34,419.57	\$784.31
	Desv. típ.	\$17,075.66	
	Varianza	291578214,453	
	Asimetría	2,125	,112
	Curtosis	5,378	,224
Meses desde el contrato	N	474	
	Rango	35	
	Mínimo	63	
	Máximo	98	
	Suma	38446	
	Media	81,11	,46
	Desv. típ.	10,06	
	Varianza	101,223	
	Asimetría	-,053	,112
	Curtosis	-1,153	,224
N válido (según	N	474	

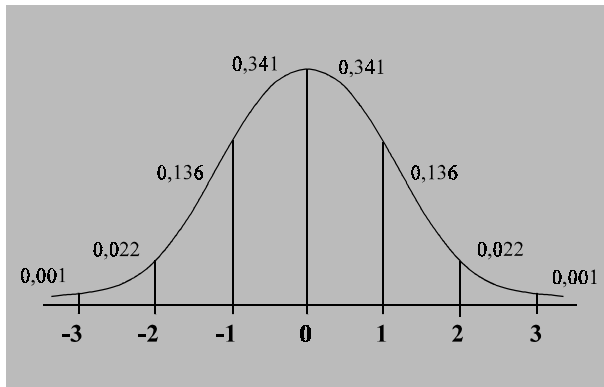
Puntuaciones típicas y curva normal

En muchas de las variables que podemos medir, la mayoría de los valores se encuentran próximos al centro de la distribución y van siendo menos frecuentes a medida que va aumentando la distancia al centro. Este tipo de distribuciones tienen forma de campana, y el ejemplo más típico es una distribución teórica llamada *curva normal*. Muchos de los fenómenos que podemos observar en la sociedad y en la naturaleza tienen distribuciones muy similares a la distribución teórica normal.

La distribución normal es, probablemente, la distribución teórica más importante en estadística y sirve como punto de referencia para describir cómo se distribuyen muchos de los datos muestrales que recogemos. La explicación de por qué esto es así se encuentra en el *teorema central del límite*, que, formulado en palabras, afirma lo siguiente: si los datos que recogemos son debidos a la suma de cierto número de causas independientes entre sí, cada una con un efecto parcial, siempre que la desviación típica de estos efectos sea finita, la distribución de los datos recogidos se asemejará tanto más a la curva normal cuantos más datos recojamos (cualquiera que sea la distribución original de esos efectos parciales).

La importancia de la distribución normal como referente del comportamiento de los datos que recogemos nos obliga a describir, aunque sólo sea brevemente, algunas de sus características (ver figura 10.8):

Figura 10.8. Curva normal (con probabilidades para algunas puntuaciones típicas).



- Tiene forma de campana (de ahí que sea conocida también como *campana de Gauss*). Esto implica que los valores centrales de la distribución son mucho más probables que los valores que se van alejando del centro de la distribución.
- Es simétrica respecto a su valor central. Al ser simétrica, las medidas de tendencia central (media, mediana, moda, etc.) coinciden.
- Es asintótica respecto al eje de abscisas (por mucho que se extienda, nunca llega a tocarlo), por lo que los valores mínimo y máximo del eje de abscisas son $-\infty$ y $+\infty$.
- Los puntos de inflexión de la curva se encuentran a una desviación típica por encima y por debajo de la media.
- Cualquier combinación lineal de variables normalmente distribuidas también se distribuye según el modelo de probabilidad normal.

La mayor parte del trabajo relacionado con variables aleatorias distribuidas normalmente consiste en hallar las probabilidades asociadas a sus valores. Estas probabilidades se obtienen integrando la función de densidad normal. Pero para evitar este tipo de cálculos se han construido tablas con las probabilidades ya calculadas. En cualquier libro de estadística puede encontrarse una tabla con las probabilidades de la curva normal. Ahora bien, esas tablas recogen las probabilidades de una curva normal muy especial: la que tiene media 0 y desviación típica 1; lo cual se expresa así: $N(0, 1)$.

Por supuesto, el hecho de que la distribución normal tabulada tenga media 0 y desviación típica 1 no es un problema sino, de hecho, una ventaja, pues cualquier variable puede ser transformada en otra variable equivalente con media 0 y desviación típica 1 sin que se alteren sus propiedades. A esta transformación la llamamos *tipificación* o *estandarización* y se realiza de la siguiente manera:

$$Z_i = \frac{X_i - E(X_i)}{S_x}$$

donde: X_i se refiere a las puntuaciones originales de la variable

$E(X_i)$ es el valor esperado de X_i (es decir, su media: \bar{X})

S_x es la desviación típica de X_i : $S_x = \left(\sum (X_i - \bar{X})^2 / n \right)^{1/2}$

Ya hemos visto que el procedimiento **Descriptivos** permite obtener puntuaciones típicas. Lógicamente, podemos transformar en puntuaciones Z_i las puntuaciones originales de cualquier variable (siempre que en esa variable tenga sentido el cálculo de la media y de la desviación típica); y una vez obtenidas las puntuaciones Z_i , podemos, por un lado, describir con precisión la posición relativa de un sujeto dentro de su distribución, pues informan sobre el número de desviaciones típicas que una determinada puntuación se aleja de su media (por arriba o por abajo, dependiendo de que la puntuación típica sea positiva o negativa, respectivamente); y, por otro, conocer la probabilidad asociada a cualquier valor X_i de una variable normalmente distribuida a partir de la probabilidad asociada a su correspondiente puntuación Z_i .

La figura 10.8 muestra las probabilidades asociadas a algunos valores Z_i (una, dos y tres desviaciones típicas por encima y por debajo del valor central). Recordemos que los valores Z_i tienen media 0 y desviación típica 1. Podemos observar en la figura 10.8 que, aunque el eje de abscisas de una curva normal admite valores desde $-\infty$ hasta $+\infty$, el 99,8 % de los casos está comprendido entre las puntuaciones 3.

Además, sabemos que en una curva normal $N(0, 1)$, entre $\pm 1,645$ puntuaciones típicas se encuentra el 90 % de los casos; entre $\pm 1,96$, el 95 % de los casos; entre $\pm 2,575$, el 99 % de los casos; etc.

< Fin del capítulo 10 >