Capítulo 17
Análisis de correlación lineal
Los procedimientos *Correlaciones bivariadas y Correlaciones parciales*

Cuando se analizan datos, el interés del analista suele centrarse en dos grandes objetivos: comparar grupos y estudiar relaciones. En el capítulo 12 hemos descrito ambos aspectos del análisis referidos a variables categóricas. En los capítulos 13, 14, 15 y 16 hemos estudiado una serie de técnicas de análisis diseñadas para comparar grupos en una variable cuantitativa. Nos falta saber cómo estudiar la relación entre variables cuantitativas.

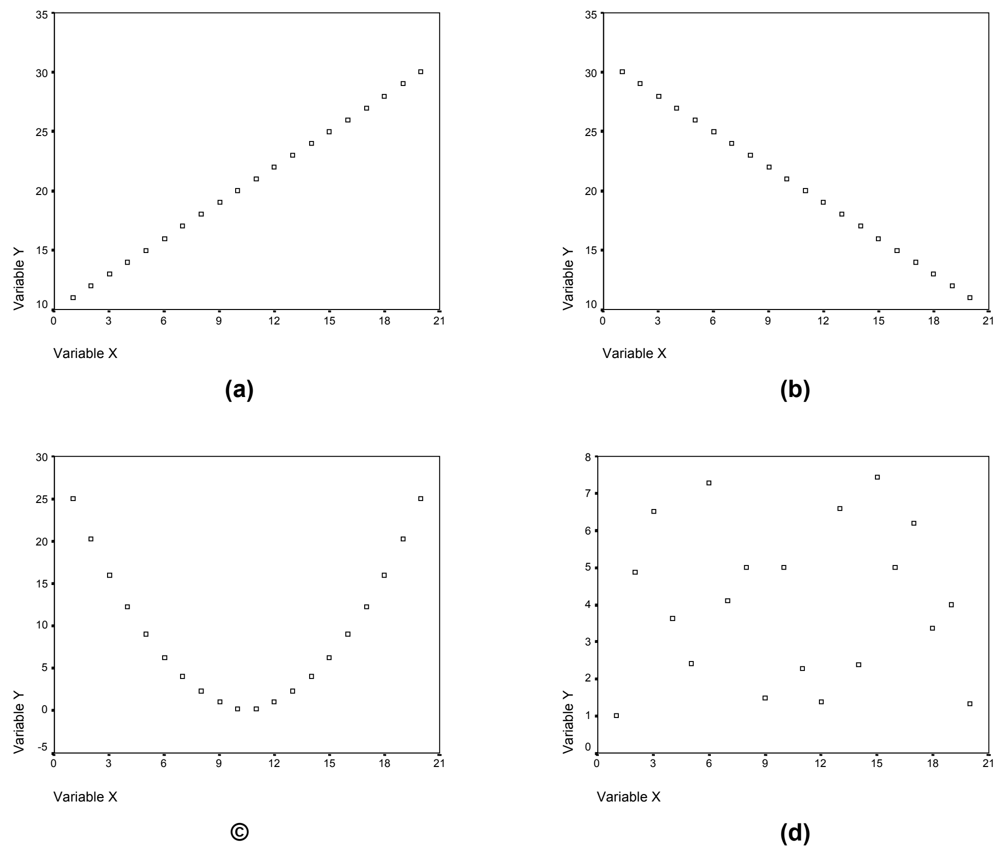
Suele decirse que los sujetos más frustrados son también más agresivos; que cuanto mayor es el nivel educativo, mayor es el nivel de renta; que los niveles altos de colesterol en sangre suelen ir acompañados de dietas alimenticias ricas en grasas; que los sujetos muestran más interés por una tarea cuanto mayor es el tamaño de la recompensa que reciben; etc. En todos los ejemplos mencionados se habla de la *relación entre dos variables*. En este capítulo se estudian algunos índices estadísticos que permiten cuantificar el grado de relación existente entre dos variables.

Correlación lineal simple

El concepto de *relación* o *correlación* se refiere al grado de variación conjunta existente entre dos o más variables. En este apartado nos vamos a centrar en el estudio de un tipo particular de relación llamada *lineal* y nos vamos limitar a considerar únicamente dos variables (*simple*). En el próximo capítulo sobre *Regresión lineal* estudiaremos el caso de más de dos variables. Una relación lineal *positiva* entre dos variables X_i e Y_i indica que los valores de las dos variables varían de forma parecida: los sujetos que puntúan alto en X_i tienden a puntuar alto en Y_i y los que puntúan bajo en X_i tienden a puntuar bajo en Y_i. Una relación lineal *negativa* significa que los valores de las dos variables varían justamente al revés: los sujetos que puntúan alto en X_i tienden a puntuar bajo en Y_i y los que puntúan bajo en X_i tienden a puntuar alto en Y_i.

La forma más directa e intuitiva de formarnos una primera impresión sobre el tipo de relación existente entre dos variables es a través de un *diagrama de dispersión*. Un diagrama de dispersión es un gráfico en el que una de las variables (X_i) se coloca en el eje de abscisas, la otra (Y_i) en el de ordenadas y los pares (x_i , y_i) se representan como una nube de puntos. La forma de la nube de puntos nos informa sobre el tipo de relación existente entre las variables. La figura 17.1 recoge cuatro diagramas de dispersión que reflejan cuatro tipos de relación diferentes.

Figura 17.1. Diagramas de dispersión expresando diferentes tipos de relación.



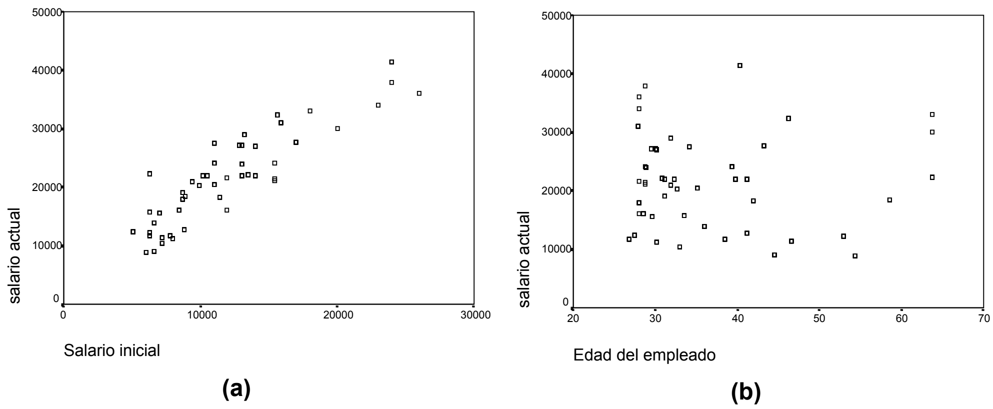
La figura 17.1.*a* muestra una situación en la que cuanto mayores son las puntuaciones en una de las variables, mayores son también las puntuaciones en la otra; cuando ocurre esto, los puntos se sitúan en una linea recta ascendente y hablamos de relación *lineal positiva*. La figura 17.1.*b* representa una situación en la que cuanto mayores son las puntuaciones en una de las variables, menores son las puntuaciones en la otra; en este caso, los puntos se sitúan en una línea recta descendente y hablamos de relación *lineal negativa*. En la situación representada en la figura 17.1.*c* también existe una pauta de variación clara, pero no es lineal: los puntos no dibujan una línea recta. Y en la figura 17.1.*d* no parece existir ninguna pauta de variación clara, lo cual queda reflejado en una nube de puntos dispersa, muy lejos de lo que podría ser una línea recta.

Vemos, pues, que un diagrama de dispersión nos permite formarnos una idea bastante aproximada sobre el *tipo de relación* existente entre dos variables. Pero, además, observando los diagramas de la figura 17.1, podemos ver que un diagrama de dispersión también puede utilizarse como una forma de *cuantificar* el grado de relación lineal existente entre dos variables: basta con observar el grado en el que la nube de puntos se ajusta a una línea recta.

Sin embargo, utilizar un diagrama de dispersión como una forma de cuantificar la relación entre dos variables no es, en la práctica, tan útil como puede parecer a primera vista. Esto es debido a que la relación entre dos variables no siempre es perfecta o nula: habitualmente no es ni lo uno ni lo otro. Consideremos los diagramas de dispersión de la figura 17.2. En el diagrama de la figura 17.2.*a*, los puntos, aun no estando situados todos ellos una línea recta, se aproximan bastante a ella. Podríamos encontrar una línea recta ascendente que representara de forma bastante aproximada el conjunto total de los puntos del diagrama, lo cual indica que la relación entre las variables *salario inicial* y *salario actual* es lineal y positiva: a mayor *salario inicial*, mayor *salario actual*.

En el diagrama 17.2.*b*, por el contrario, da la impresión de que no hay forma de encontrar una recta a la que poder aproximar los puntos. Al margen de que entre las variables *edad* y *salario actual* pueda existir algún tipo de relación, parece claro que la relación no es de tipo lineal.

Figura 17.2. Diagramas de dispersión representando *relación lineal* (a) e *independencia lineal* (b).



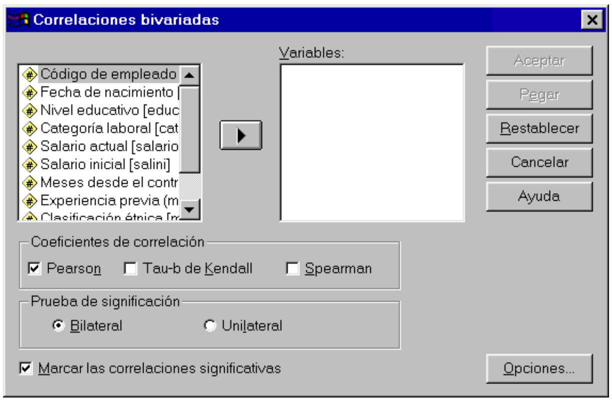
Estas consideraciones sugieren que hay nubes de puntos a las que es posible ajustar una línea recta mejor de lo que es posible hacerlo a otras. Por lo que el ajuste de una recta a una nube de puntos no parece una cuestión de todo o nada, sino más bien de *grado* (más o menos ajuste). Lo cual nos advierte sobre la necesidad de utilizar algún índice numérico capaz de cuantificar ese grado de ajuste con mayor precisión de lo que nos permite hacerlo una simple inspección del diagrama de dispersión.

Estos índices numéricos suelen denominarse *coeficientes de correlación* y poseen la importante propiedad de permitir cuantificar el grado de relación lineal existente entre dos variables cuantitativas. Por supuesto, al mismo tiempo que permiten cuantificar el grado de relación lineal existente entre dos variables, también sirven para valorar el grado de ajuste de la nube de puntos a una línea recta.

Para obtener algunos de estos coeficientes de correlación:

▶ Seleccionar la opción **Correlaciones** > **Bivariadas** del menú **Analizar** para acceder al cuadro de diálogo *Correlaciones Bivariadas* que muestra la figura 17.3.

Figura 17.3. Cuadro de diálogo *Correlaciones bivariadas*.



La lista de variables muestra únicamente las variables del archivo de datos que poseen formato numérico. Desde este cuadro de diálogo es posible obtener varios coeficientes de correlación y algunos estadísticos descriptivos básicos. Para ello:

- Seleccionar las variables cuantitativas cuyo grado de relación se desea estudiar y trasladarlas a la lista **Variables**. Es necesario trasladar al menos dos variables.

Coeficientes de correlación. Pueden seleccionarse uno o más de los siguientes tres coeficientes de correlación:

- **Pearson**. El coeficiente de correlación de Pearson (1896) es, quizá, el mejor coeficiente y el más utilizado para estudiar el grado de relación lineal existente entre dos variables cuantitativas. Se suele representar por r y se obtiene tipificando el promedio de los productos de las puntuaciones diferenciales de cada caso (desviaciones de la media) en las dos variables correlacionadas:

$$r_{xy} = \frac{\sum x_i y_i}{n S_x S_y}$$

(x_i e y_i se refieren a las puntuaciones diferenciales de cada par; n al número de casos; y S_x y S_y a las desviaciones típicas de cada variable).

El coeficiente de correlación de Pearson toma valores entre -1 y 1: un valor de 1 indica relación lineal perfecta positiva; un valor de -1 indica relación lineal perfecta negativa (en ambos casos los puntos se encuentran dispuestos en una línea recta); un valor de 0 indica relación lineal nula (lo que ocurre, por ejemplo, en los ejemplos de las figuras 17.1.*c* y 17.1.*d*). El coeficiente r es una medida simétrica: la correlación entre X_i e Y_i es la misma que entre Y_i y X_i.

Es importante señalar que un coeficiente de correlación alto no implica *causalidad*. Dos variables pueden estar linealmente relacionadas (incluso muy relacionadas) sin que una sea causa de la otra.

Al marcar la opción **Pearson** el *Visor* ofrece una matriz de correlaciones cuadrada, con *unos* en la diagonal (pues la relación entre una variable y ella misma es perfecta —si bien esos *unos* son el resultado de tipificar la varianza de cada variable) y con los coeficientes de correlación entre cada dos variables duplicados en los triángulos superior e inferior de la matriz. Cada coeficiente aparece acompañado del número de casos sobre el que ha sido calculado y del nivel crítico que le corresponde bajo la hipótesis nula de que su verdadero valor poblacional es cero.

☐ **Tau-b de Kendall**. Este coeficiente de correlación es apropiado para estudiar la relación entre variables ordinales. Se basa en el número de inversiones y no inversiones entre casos y ya ha sido descrito en el capítulo 12, en el apartado *Estadísticos: Datos ordinales*. Toma valores entre −1 y 1, y se interpreta exactamente igual que el coeficiente de correlación de Pearson.

La utilización de este coeficiente tiene sentido si las variables no alcanzan el nivel de medida de intervalo y/o no podemos suponer que la distribución poblacional conjunta de las variables sea normal.

☐ **Spearman**. El coeficiente de correlación *rho* de Spearman (1904) es el coeficiente de correlación de Pearson, pero aplicado después de transformar las puntuaciones originales en rangos. Toma valores entre −1 y 1, y se interpreta exactamente igual que el coeficiente de correlación de Pearson.

Al igual que ocurre con el coeficiente *tau-b* de Kendall, el de Spearman puede utilizarse como una alternativa al de Pearson cuando las variables estudiadas son ordinales y/o se incumple el supuesto de normalidad.

Prueba de significación. Junto con cada coeficiente de correlación, el *Visor* ofrece la información necesaria para contrastar la hipótesis nula de que el valor poblacional del coeficiente es cero. Esta hipótesis se contrasta mediante un valor tipificado que, en el caso del coeficiente de correlación de Pearson, adopta la siguiente forma:

$$T = r_{xy}\sqrt{n-2}\Big/\sqrt{1-r_{xy}^2}$$

Si suponemos que la muestra utilizada ha sido aleatoriamente extraída de una población en la que las dos variables correlacionadas se distribuyen normalmente, el estadístico T se distribuye según el modelo de probabilidad t de Student con $n-2$ grados de libertad. El SPSS permite seleccionar el nivel crítico deseado:

○ **Bilateral**. Opción apropiada para cuando no existen expectativas sobre la *dirección* de la relación. Indica la probabilidad de obtener coeficientes tan alejados de cero o más que el valor obtenido.

○ **Unilateral**. Opción apropiada para cuando existen expectativas sobre la *dirección* de la relación. Indica la probabilidad de obtener coeficientes tan grandes o más grandes que el obtenido si el coeficiente es positivo, o tan pequeños o más pequeños que el obtenido si el coeficiente es negativo.

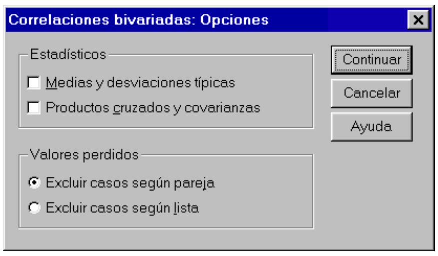
☐ **Marcar las correlaciones significativas**. Esta opción, que se encuentra activa por defecto, permite obtener el nivel crítico exacto asociado a cada coeficiente de correlación. Si se desactiva esta opción, en lugar del nivel crítico, el *Visor* muestra un asterisco al lado de los coeficientes con nivel crítico menor que 0,05 y dos asteriscos al lado de los coeficientes con nivel crítico menor que 0,01.

Opciones

Para obtener alguna información adicional (algunos estadísticos descriptivos, la covarianza, etc.) y controlar el tratamiento que se desea dar a los valores perdidos:

▶ Pulsar el botón **Opciones...** del cuadro de diálogo *Correlaciones bivariadas* (ver figura 17.3) para acceder al subcuadro de diálogo *Correlaciones bivariadas: Opciones* que muestra la figura 17.4.

Figura 17.4. Subcuadro de diálogo *Correlaciones bivariadas: Opciones*.



Estadísticos. Si se ha elegido el coeficiente de correlación de Pearson (ver figura 17.3), este recuadro permite seleccionar una o más de las siguientes opciones:

☐ **Medias y desviaciones típicas**. Muestra, para cada variable, la media aritmética, la desviación típica insesgada y el número de casos válidos.

☐ **Productos cruzados y covarianzas**. Muestra, para cada par de variables, el numerador del coeficiente de correlación de Pearson (es decir, los productos cruzados de las desviaciones de cada puntuación respecto de su media) y ese mismo numerador dividido por $n-1$ (es decir, la covarianza).

Valores perdidos. Las dos opciones de este recuadro permiten seleccionar el tratamiento que se desea dar a los valores perdidos.

○ **Excluir casos según pareja**. Se excluyen del cálculo de cada coeficiente de correlación los casos con valor perdido en alguna de las dos variables que se están correlacionando.

○ **Excluir casos según lista**. Se excluyen del cálculo de todos los coeficientes de correlación solicitados los casos con valor perdido en cualquiera de las variables seleccionadas en la lista **Variables**.

Ejemplo (Correlaciones > Bivariadas)

Este ejemplo muestra cómo obtener los coeficientes de correlación y los estadísticos del procedimiento **Correlaciones bivariadas**.

- En el cuadro de diálogo *Correlaciones bivariadas* (ver figura 17.3), seleccionar las variables *tiempemp* (meses desde el contrato), *salini* (salario inicial) y *salario* (salario actual) y trasladarlas a la lista **Variables**.
- Marcar las opciones **Pearson**, **Tau-b de Kendall** y **Spearman** del recuadro **Coeficientes de correlación**.
- Pulsar el botón **Opciones...** para acceder al cuadro de diálogo *Correlaciones bivariadas: Opciones* (ver figura 17.4) y, en el recuadro **Estadísticos**, marcar las opciones **Medias y desviaciones típicas** y **Productos cruzados y covarianzas**.

Aceptando estas elecciones, el *Visor de resultados* ofrece la información que recogen las tablas 17.1, 17.2 y 17.3. La primera de ellas (tabla 17.1) contiene información descriptiva: la media aritmética, la desviación típica insesgada y el número de casos válidos; todo ello, para cada variable individualmente considerada.

Tabla 17.1. Tabla de *estadísticos descriptivos*.

	Media	Desviación típica	N
Meses desde el contrato	81,11	10,06	474
Salario inicial	$17,016.09	$7,870.64	474
Salario actual	$34,419.57	$17,075.66	474

Tabla 17.2. Tabla resumen del *coeficiente de correlación de Pearson*.

		Meses desde el contrato	Salario inicial	Salario actual
Meses desde el contrato	Correlación de Pearson	1,000	-,020	,084
	Sig. (bilateral)	,	,668	,067
	Suma de cuadr. y prod. cruzados	47878,295	-739866,498	6833347,489
	Covarianza	101,223	-1564,200	14446,823
	N	474	474	474
Salario inicial	Correlación de Pearson	-,020	1,000	,880**
	Sig. (bilateral)	,668	,	,000
	Suma de cuadr. y prod. cruzados	-739866,498	29300904965,454	55948605047,732
	Covarianza	-1564,200	61946944,959	118284577,268
	N	474	474	474
Salario actual	Correlación de Pearson	,084	,880**	1,000
	Sig. (bilateral)	,067	,000	,
	Suma de cuadr. y prod. cruzados	6833347,489	55948605047,732	137916495436,34
	Covarianza	14446,823	118284577,268	291578214,453
	N	474	474	474

**. La correlación es significativa al nivel 0,01 (bilateral).

La tabla 17.2 ofrece la información referida al *coeficiente de correlación de Pearson*. Cada celda contiene cinco valores referidos al cruce entre cada dos variables: 1) el valor del coeficiente de correlación de Pearson; 2) el nivel crítico bilateral que corresponde a ese coeficiente (*Sig. bilateral*; el nivel crítico unilateral puede obtenerse dividiendo por 2 el bilateral); 3) la suma de cuadrados (para el cruce de una variable consigo misma) y la suma de productos cruzados (para el cruce de dos variables distintas); 4) la covarianza; y 5) el número de casos válidos (*N*) sobre el que se han efectuado los cálculos.

El nivel crítico permite decidir sobre la hipótesis nula de independencia lineal (o lo que es lo mismo, sobre la hipótesis de que el coeficiente de correlación vale cero en la población). Rechazaremos la hipótesis nula de independencia (y concluiremos que existe relación lineal significativa) cuando el nivel crítico sea menor que el nivel de significación establecido (generalmente, 0,05). Así, basándonos en los niveles críticos de la tabla 17.2, podemos afirmar que, mientras las variables *salario inicial* y *salario actual* correlacionan significativamente (*Sig.* = 0,000), la variable *meses desde el contrato* no correlaciona ni con la variable *salario inicial* (*Sig.* = 0,668) ni con la variable *salario actual* (*Sig.* = 0,067).

El SPSS no puede calcular un coeficiente de correlación cuando todos los casos de una de las variables (o de las dos) son casos con valores perdidos, o cuando todos los casos tienen el mismo valor en una o en las dos variables correlacionadas. Cuando ocurre esto, el SPSS sustituye el coeficiente de correlación por una coma (también muestra una coma en lugar del nivel crítico —*Sig.*—que corresponde al cruce de una variable consigo misma).

Tabla 17.3. Tabla resumen: coeficientes *rho de Spearman* y *tau-b de Kendall*.

			Meses desde el contrato	Salario inicial	Salario actual
Tau_b de Kendall	Meses desde el contrato	Coeficiente de correlación	1,000	-,046	,071*
		Sig. (bilateral)	,	,146	,022
		N	474	474	474
	Salario inicial	Coeficiente de correlación	-,046	1,000	,656**
		Sig. (bilateral)	,146	,	,000
		N	474	474	474
	Salario actual	Coeficiente de correlación	,071*	,656**	1,000
		Sig. (bilateral)	,022	,000	,
		N	474	474	474
Rho de Spearman	Meses desde el contrato	Coeficiente de correlación	1,000	-,063	,105*
		Sig. (bilateral)	,	,168	,023
		N	474	474	474
	Salario inicial	Coeficiente de correlación	-,063	1,000	,826**
		Sig. (bilateral)	,168	,	,000
		N	474	474	474
	Salario actual	Coeficiente de correlación	,105*	,826**	1,000
		Sig. (bilateral)	,023	,000	,
		N	474	474	474

*. La correlación es significativa al nivel 0,05 (bilateral).

**. La correlación es significativa al nivel 0,01 (bilateral).

La tabla 17.3. recoge la información referida a los coeficientes *tau-b de Kendall* y *rho de Spearman*. En esta tabla aparecen tres valores por cada cruce de variables: 1) el valor del coeficiente de correlación; 2) el nivel crítico asociado a cada coeficiente (*Sig.*); y 3) el número de casos sobre el que se ha calculado cada coeficiente. Puesto que estos coeficientes se basan en las propiedades ordinales de los datos, su valor y su nivel crítico no tienen por qué ser los mismos que los obtenidos con el coeficiente de correlación de Pearson. De hecho, tanto con el coeficiente *tau-b* como con el coeficiente *rho*, la relación entre las variables *meses desde el contrato* y *salini* ha pasado a ser significativa (0,022).

Correlación parcial

El procedimiento **Correlaciones parciales** permite estudiar la relación lineal existente entre dos variables controlando el posible efecto de una o más variables extrañas. Un coeficiente de *correlación parcial* es una técnica de control estadístico que expresa el grado de relación lineal existente entre dos variables tras eliminar de ambas el efecto atribuible a terceras variables.

Por ejemplo, se sabe que la correlación entre las variables *inteligencia* y *rendimiento escolar* es alta y positiva. Sin embargo, cuando se controla el efecto de terceras variables como el *número de horas de estudio* o el *nivel educativo de los padres*, la correlación entre *inteligencia* y *rendimiento* desciende, lo cual indica que la relación entre inteligencia y rendimiento está condicionada, depende o está modulada por las variables sometidas a control.

Para obtener correlaciones parciales:

- Seleccionar la opción **Correlaciones** > **Parciales...** del menú **Analizar** para acceder al cuadro de diálogo *Correlaciones parciales* que muestra la figura 17.5.

Figura 17.5. Cuadro de diálogo *Correlaciones parciales*.

La lista de variables muestra un listado de las variables del archivo de datos que poseen formato numérico. Para obtener un coeficiente de correlación parcial:

- Trasladar a la lista **Variables** las variables que interesa correlacionar.
- Trasladar a la lista **Controlando para** las variables cuyo efecto se desea controlar.

El procedimiento **Correlaciones parciales** puede manipular un total de 400 variables, de las cuales hasta un máximo de 100 pueden ser variables sobre las que se ejerce control.

La ecuación para obtener el coeficiente de correlación parcial depende del número de variables que se estén controlando:

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \qquad \textbf{(correlación parcial de primer orden)}$$

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}} \qquad \textbf{(correlación parcial de segundo orden)}$$

Los coeficientes de mayor orden se obtienen siguiendo la misma lógica. Hablamos de correlación de *primer orden* para indicar que se está controlando el efecto de una variable; de *segundo orden*, para indicar que se está controlando el efecto de dos variables; etc. Lógicamente, cuando no se está controlando ninguna variable, es decir, cuando utilizamos el coeficiente de correlación de Pearson del apartado anterior, hablamos de correlación de *orden cero*.

Prueba de significación. Junto con cada coeficiente de correlación parcial, el *Visor* ofrece la información necesaria para contrastar la hipótesis nula de que el valor poblacional del coeficiente de correlación vale cero. Esta hipótesis se contrasta mediante un valor tipificado del coeficiente de correlación parcial que adopta la siguiente forma:

$$T = r_{12.k}\sqrt{m-k-2} \Big/ \sqrt{1-r_{12.k}^2}$$

donde m se refiere al número mínimo de casos con puntuación válida en el conjunto de posibles correlaciones de orden cero entre cada par de variables seleccionadas (es decir, el número de casos de la correlación de orden cero con menor número de casos válidos) y k es el número de variables controladas.

El estadístico T permite contrastar la hipótesis nula de que el valor poblacional del coeficiente de correlación parcial es cero. Este estadístico se distribuye según el modelo de probabilidad t de Student con $m-k-2$ grados de libertad. El SPSS permite seleccionar el tipo de nivel crítico deseado:

- ○ **Bilateral**. Opción apropiada para cuando no existen expectativas sobre la *dirección* de la relación. Indica la probabilidad de obtener coeficientes tan alejados de cero o más que el valor absoluto del coeficiente obtenido.
- ○ **Unilateral**. Opción apropiada para cuando existen expectativas sobre la *dirección* de la relación. Indica la probabilidad de obtener coeficientes tan grandes o más grandes que el obtenido si el coeficiente es positivo, o tan pequeños o más pequeños que el obtenido si el coeficiente de correlación es negativo.

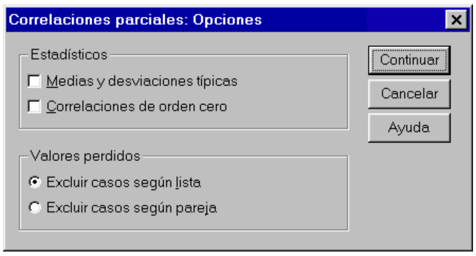
☐ **Mostrar el nivel de significación real**. Esta opción, que se encuentra activa por defecto, permite obtener el nivel crítico exacto y los grados de libertad asociados a cada coeficiente de correlación parcial. Al desactivar esta opción, en lugar del nivel crítico exacto, el *Visor* muestra un asterisco al lado de los coeficientes con nivel crítico menor o igual que 0,05 y dos asteriscos al lado de los coeficientes con nivel crítico menor o igual que 0,01.

Opciones

Para obtener alguna información adicional (algunos estadísticos descriptivos y los coeficientes de correlación de orden cero) y para controlar el tratamiento que se desea dar a los valores perdidos:

▶ Pulsar el botón **Opciones...** del cuadro de diálogo *Correlaciones parciales* (ver figura 17.5) para acceder al subcuadro de diálogo *Correlaciones parciales: Opciones* que muestra la figura 17.6.

Figura 17.6. Subcuadro de diálogo *Correlaciones parciales: Opciones*.



Estadísticos. Este recuadro permite seleccionar una o más de las siguientes opciones:

☐ **Medias y desviaciones típicas**. Para obtener la media aritmética, la desviación típica insesgada y el número de casos válidos de cada variable individualmente considerada.

☐ **Correlaciones de orden cero**. Para obtener los coeficientes de correlación de orden cero entre cada para de variables (es decir, para obtener el coeficiente de correlación de Pearson entre cada par de variables sin ejercer ningún control sobre terceras variables).

Valores perdidos. Las dos opciones de este recuadro permiten seleccionar el tratamiento que se desea dar a los valores perdidos.

☐ **Excluir casos según pareja**. Se excluyen del cálculo de cada coeficiente de correlación los casos con valor perdido en alguna de las variables que están interviniendo en el coeficiente de correlación parcial.

☐ **Excluir casos según lista**. Se excluyen del cálculo de todos los coeficientes de correlación solicitados los casos con valor perdido en cualquiera de las variables seleccionadas en la lista **Variables**.

Ejemplo (Correlaciones > parciales)

Este ejemplo muestra cómo utilizar el procedimiento **Correlaciones parciales** para estudiar la relación entre dos variables cuando se controla el efecto de terceras variables:

- ▶ En el cuadro de diálogo *Correlaciones parciales* (ver figura 17.5), seleccionar las variables *salini* (salario inicial) y *salario* (salario actual) y trasladarlas a la lista **Variables**.
- ▶ Seleccionar las variables *educ* (nivel educativo), *tiempemp* (meses desde el contrato) y *expprev* (experiencia previa) y trasladarlas a la lista **Controlando para**.
- ▶ Pulsar el botón **Opciones...** para acceder al cuadro de diálogo *Correlaciones parciales: Opciones* (ver figura 17.6) y, en el recuadro **Estadísticos**, marcar las opciones **Medias y desviaciones típicas** y **Correlaciones de orden cero**.

Aceptando estas elecciones, el *Visor* ofrece los resultados que aparecen a continuación (los resultados de este procedimiento no aparecen en formato de tabla pivotante). El primer bloque de información ofrece la media aritmética, la desviación típica insesgada y el número de casos válidos; todo ello, para cada variable individualmente considerada.

Variable	Mean	Standard Dev	Cases
SALARIO	34419,5675	17075,6615	474
SALINI	17016,0865	7870,6382	474
EDUC	13,4916	2,8848	474
TIEMPEMP	81,1097	10,0609	474
EXPPREV	95,8608	104,5862	474

A continuación aparece una matriz con los coeficientes de correlación de orden cero (sin parcializar efectos) entre todas las variables seleccionadas:

```
Zero Order Partials

              SALARIO     SALINI       EDUC   TIEMPEMP    EXPPREV

SALARIO        1,0000      ,8801      ,6606      ,0841     -,0975
              (    0)    (  472)    (  472)    (  472)    (  472)
              P= ,       P= ,000    P= ,000    P= ,067    P= ,034

SALINI          ,8801     1,0000      ,6332     -,0198      ,0451
              (  472)    (    0)    (  472)    (  472)    (  472)
              P= ,000    P= ,       P= ,000    P= ,668    P= ,327

EDUC            ,6606      ,6332     1,0000      ,0474     -,2524
              (  472)    (  472)    (    0)    (  472)    (  472)
              P= ,000    P= ,000    P= ,       P= ,303    P= ,000

TIEMPEMP        ,0841     -,0198      ,0474     1,0000      ,0030
              (  472)    (  472)    (  472)    (    0)    (  472)
              P= ,067    P= ,668    P= ,303    P= ,       P= ,948

EXPPREV        -,0975      ,0451     -,2524      ,0030     1,0000
              (  472)    (  472)    (  472)    (  472)    (    0)
              P= ,034    P= ,327    P= ,000    P= ,948    P= ,

(Coefficient / (D.F.) / 2-tailed Significance)
```

La matriz muestra, para cada par de variables, el coeficiente de correlación de Pearson, los grados de libertad asociados al estadístico de contraste *T* (número de casos válidos menos dos), y el nivel crítico bilateral (el unilateral se obtiene dividiendo el bilateral por dos). La información de esta matriz es doblemente útil: por un lado, informa sobre el grado de relación existente entre las dos variables que interesa estudiar (en nuestro ejemplo, *salini* y *salario*); por otro, permite averiguar si las variables cuyo efecto se desea controlar (*educ*, *tiempemp* y *expprev*) están o no relacionas con las dos variables que interesa correlacionar (*salini* y *salario*).

Así, podemos ver que el coeficiente de correlación entre *salini* y *salario* vale 0,88, con un nivel crítico $p = 0{,}000$ que nos permite afirmar que el coeficiente es significativamente distinto de cero. También podemos ver que, de las tres variables control utilizadas, *educ* correlaciona significativamente tanto con *salini* ($p = 0{,}000$) como con *salario* ($p = 0{,}000$), *tiempemp* no correlaciona ni con *salini* ($p = 0{,}668$) ni con *salario* ($p = 0{,}067$), y *expprev* correlaciona con *salario* ($p = 0{,}000$), pero no con *salini* ($p = 0{,}000$).

El último bloque de información ofrece el coeficiente de correlación parcial entre *salini* y *salario*:

```
Controlling for..    EDUC      TIEMPEMP  EXPPREV

             SALARIO     SALINI

SALARIO       1,0000      ,8118
             (     0)    (  469)
             P= ,        P= ,000

SALINI         ,8118     1,0000
             (  469)    (     0)
             P= ,000    P= ,

(Coefficient / (D.F.) / 2-tailed Significance)
```

El coeficiente de correlación parcial entre las variables *salini* y *salario* (es decir, el coeficiente de correlación obtenido tras eliminar de ambas variables el efecto de las variables *educ*, *tiempemp* y *expprev*) vale 0,81, con un nivel crítico $p = 0{,}000$ que nos permite afirmar que es significativamente distinto de cero. Puesto que el coeficiente de correlación parcial permanece significativo y su diferencia con el coeficiente de orden cero es más bien escasa (ha bajado de 0,88 a 0,81), podemos afirmar: 1) que entre las variables *salini* y *salario* existe relación lineal significativa, y 2) que tal relación no se ve sustancialmente alterada tras controlar el efecto de las variables *educ*, *tiempemp* y *expprev*.