

Capítulo 13

Contrastes sobre medias: Los procedimientos *Medias* y *Prueba T*

La opción **Comparar medias** del menú **Analizar** contiene varios de los procedimientos estadísticos diseñados para efectuar contrastes de hipótesis sobre medias; en concreto, la prueba *T* y el análisis de varianza de un factor.

En este capítulo vamos a estudiar cuatro procedimientos SPSS: *Medias*, *Prueba T para una muestra*, *Prueba T para dos muestras independientes* y *Prueba T para dos muestras relacionadas*. El próximo capítulo estará dedicado por completo al *Análisis de varianza de un factor*.

Medias

El procedimiento **Medias** ofrece, como utilidad fundamental, estadísticos descriptivos que pueden calcularse teniendo en cuenta los distintos grupos y subgrupos definidos por una o más variables independientes. Opcionalmente, pueden solicitarse ANOVAs de un factor, obtener algunos estadísticos sobre proporción de varianza explicada y contrastar la hipótesis de linealidad (si bien estas opciones se estudiarán detalladamente en el próximo capítulo). Para utilizar el procedimiento **Medias**:

- ▣ Seleccionar la opción **Comparar medias > Medias...** del menú **Analizar**. Esta acción conduce al cuadro de diálogo *Medias* que muestra la figura 13.1

Figura 13.1. Cuadro de diálogo *Medias*.

Para obtener los estadísticos descriptivos que el procedimiento ofrece por defecto (la *media aritmética*, la *desviación típica* y el *número de casos* de cada variable *dependiente* en cada uno de los grupos definidos por cada variable *independiente*):

- ▶ Trasladar a la lista **Dependientes** la variable o variables que interesa describir o aquellas en las que se van a comparar los grupos.
- ▶ Trasladar a la lista **Independientes** la variable o variables que definen los grupos que interesa describir o comparar.
- ▶ Pulsar el botón **Aceptar**.

Con estas especificaciones mínimas, el SPSS ofrece información descriptiva para cada uno de los grupos definidos por los niveles de las variables independientes seleccionadas, es decir, sin combinar entre sí los niveles de las distintas variables independientes. Pero también es posible combinar más de una variable independiente para, dentro de los grupos definidos por una primera variable, formar subgrupos definidos por una segunda variable (o una tercera, o una cuarta, etc.). Es decir, cada estadístico solicitado puede calcularse en cada uno de los subgrupos resultantes de combinar dos o más variables independientes. Esto se consigue definiendo *capas*. Para definir capas:

- ▶ Utilizar los botones **Siguiente** y **Anterior** del apartado **Capa # de #** (ver figura 13.1). El primer símbolo # se refiere al número de capa en el que nos encontramos; el segundo símbolo # se refiere al número de capas definidas.

Para obtener, por ejemplo, estadísticos descriptivos de la variable *salario* (salario actual) en cada uno de los subgrupos resultantes de combinar las variables *sexo* y *catlab* (categoría laboral):

- ▶ Seleccionar la variable *salario* como variable **Dependiente**.
- ▶ Seleccionar la variable *sexo* como variable **Independiente** en la *primera capa*.
- ▶ Pulsar el botón **Siguiente**.
- ▶ Seleccionar la variable *catlab* como variable **Independiente** en la *segunda capa*.
- ▶ Utilizar el botón **Anterior** para moverse por capas previamente definidas.

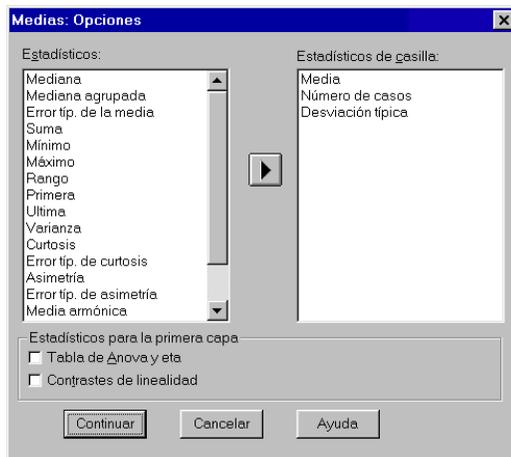
Con estas elecciones, el *Visor de resultados* ofrece algunos estadísticos descriptivos de la variable *salario* para cada uno de los 6 grupos resultantes de combinar los 3 niveles de la variable *catlab* con los 2 de la variable *sexo*. Conforme vamos creando capas, los valores # del recuadro **Capa # de #** van indicando el número de la capa en el que nos encontramos y el número total de capas definidas.

Opciones

Las opciones del procedimiento **Medias** permiten seleccionar: 1) los estadísticos descriptivos concretos que interesa obtener, y 2) algunos contrastes (sobre medias, sobre linealidad) que el procedimiento no lleva a cabo por defecto. Para solicitar estos estadísticos y contrastes:

- ▶ Pulsar el botón **Opciones...** del cuadro de diálogo *Medias* (ver figura 13.1) para acceder al subcuadro de diálogo *Medias: Opciones* que muestra la figura 13.2.

Figura 13.2. Subcuadro de diálogo *Medias: Opciones*.



Estadísticos. Contiene todos los estadísticos disponibles en el procedimiento **Medias**.

Estadísticos de casilla. Contiene los estadísticos que el procedimiento **Medias** calcula por defecto. Cualquier estadístico adicional que se desee obtener debe trasladarse a esta lista desde la lista **Estadísticos**.

Estadísticos para la primera capa. Para los grupos definidos por las variables independientes seleccionadas en la primera capa (sólo en la primera capa), es posible marcar una o más de las siguientes opciones (ambas se describen con detalle en el próximo capítulo):

- **Tabla de ANOVA y eta.** Ofrece la tabla resumen del análisis de varianza de un factor y algunos estadísticos sobre la proporción de varianza explicada: el coeficiente de correlación de Pearson y su cuadrado (sólo si el formato de la variable independiente no es de *cadena*), y el coeficiente de correlación eta y su cuadrado.
- **Contrastes de linealidad.** Los contrastes de linealidad permiten averiguar si una variable independiente *categorica* se relaciona linealmente con una variable dependiente *cuantitativa*. Los contrastes de linealidad no están disponibles para las variables independientes con formato de cadena.

Ejemplo (Comparar medias > Medias)

Este ejemplo muestra cómo utilizar el procedimiento **Medias** para obtener algunos estadísticos de la variable *salario* en los subgrupos definidos por las variables *sexo* y *catlab*:

- ▶ En el cuadro de diálogo *Medias* (ver figura 13.1), seleccionar la variable *salario* (salario actual) como **Dependiente** y la variable *sexo* como **Independiente**.
- ▶ Pulsar el botón **Siguiente** del recuadro **Capa 1 de 1** y seleccionar la variable *catlab* (categoría laboral) como variable **Independiente** en la *segunda capa*.
- ▶ Pulsar el botón **Opciones...** para acceder al subcuadro de diálogo *Medias: Opciones* (ver figura 13.2), seleccionar todos los estadísticos de la lista **Estadísticos** y trasladarlos a la lista **Estadísticos de casilla** (no marcar las opciones **Tabla de ANOVA y eta** y **Contrastes de linealidad**; estas opciones se describen con detalle en el próximo capítulo).

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestra la tabla 13.1 (hemos pivotado la tabla para adaptarla al tamaño de la página).

Tabla 13.1. Tabla de *Estadísticos* del procedimiento *Medias*.

Salario actual

Sexo		Categoría laboral			
		Administrativo	Seguridad	Directivo	Total
Hombre	N	157	27	74	258
	Media	\$31,558.15	\$30,938.89	\$66,243.24	\$41,441.78
	Desv. típ.	\$7,997.98	\$2,114.62	\$18,051.57	\$19,499.21
	Mediana	\$29,850.00	\$30,750.00	\$63,750.00	\$32,850.00
	Mediana agrupada	\$29,737.50	\$30,725.00	\$63,125.00	\$32,850.00
	Error típ. de la media	\$638.31	\$406.96	\$2,098.45	\$1,213.97
	Suma	\$4,954,630	\$835,350	\$4,902,000	\$10,691,980
	Mínimo	\$19,650	\$24,300	\$38,700	\$19,650
	Máximo	\$80,000	\$35,250	\$135,000	\$135,000
	Rango	\$60,350	\$10,950	\$96,300	\$115,350
	Varianza	63967646,887	4471602,564	325859166,050	380219336,303
	Curtosis	9,850	3,652	2,116	2,780
	Error típ. de la curtosis	,385	,872	,552	,302
	Asimetría	2,346	-,368	1,193	1,639
	Error típ. de la asimetría	,194	,448	,279	,152
	Media armónica	\$30,070.04	\$30,793.00	\$62,096.13	\$35,392.56
	Media geométrica	\$30,750.38	\$30,867.29	\$64,071.30	\$37,972.18
	% de la suma total	30,4%	5,1%	30,0%	65,5%
	% del total de N	33,1%	5,7%	15,6%	54,4%
Mujer	N	206		10	216
	Media	\$25,003.69		\$47,213.50	\$26,031.92
	Desv. típ.	\$5,812.84		\$8,501.25	\$7,558.02
	Mediana	\$24,000.00		\$45,187.50	\$24,300.00
	Mediana agrupada	\$24,066.67		\$45,187.50	\$24,275.00
	Error típ. de la media	\$405.00		\$2,688.33	\$514.26
	Suma	\$5,150,760		\$472,135	\$5,622,895
	Mínimo	\$15,750		\$34,410	\$15,750
	Máximo	\$54,000		\$58,125	\$58,125
	Rango	\$38,250		\$23,715	\$42,375
	Varianza	33789086,810		72271294,722	57123688,268
	Curtosis	4,029		-1,554	4,641
	Error típ. de la curtosis	,337		1,334	,330
	Asimetría	1,421		-,019	1,863
	Error típ. de la asimetría	,169		,687	,166
	Media armónica	\$23,861.66		\$45,805.40	\$24,402.89
	Media geométrica	\$24,406.44		\$46,511.67	\$25,146.07
	% de la suma total	31,6%		2,9%	34,5%
	% del total de N	43,5%		2,1%	45,6%

Prueba T para una muestra

La prueba T para una muestra permite contrastar hipótesis referidas a una media poblacional. Si de una población en la que la variable Y_i se distribuye normalmente con media μ seleccionamos aleatoriamente una muestra de tamaño n , podemos tipificar la media \bar{Y} de esa muestra restándole su valor esperado (que es justamente la media de la población) y dividiendo esa diferencia por su error típico (que podemos llamar $\sigma_{\bar{Y}}$), es decir:

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}}$$

Obtenemos así una puntuación Z normalmente distribuida con media 0 y desviación típica 1, que puede ser interpretada utilizando las probabilidades de la distribución normal estandarizada: sabemos, por ejemplo, que entre 1,96 puntuaciones típicas se encuentra el 95 % de los casos; que entre $\pm 2,58$ puntuaciones típicas se encuentra el 99 % de los casos, etc.

El error típico de la media ($\sigma_{\bar{Y}}$) es la desviación típica de la distribución muestral de \bar{Y} , es decir, la desviación típica de las medias calculadas en todas las muestras de tamaño n que es posible extraer de una determinada población. Se obtiene mediante: $\sigma_{\bar{Y}} = \sigma / \sqrt{n}$ (siendo σ la desviación típica de la población). El problema con el que nos encontramos al calcular $\sigma_{\bar{Y}}$ es que el valor de la desviación típica poblacional σ es, generalmente, desconocido. Necesitamos estimarlo utilizando la desviación típica (insesgada) muestral, S_{n-1} , en cuyo caso, el error típico de la media se obtiene mediante: $\hat{\sigma}_{\bar{Y}} = S_{n-1} / \sqrt{n}$.

El hecho de tener que estimar la desviación típica poblacional hace que la tipificación del estadístico \bar{Y} ya no sea una puntuación Z , sino una puntuación T :

$$T = \frac{\bar{Y} - \mu}{\hat{\sigma}_{\bar{Y}}} = \frac{\bar{Y} - \mu}{S_{n-1} / \sqrt{n}}$$

distribuida según el modelo t de *Student* con $n-1$ grados de libertad. Esta tipificación del estadístico \bar{Y} es lo que se conoce como *prueba T para una muestra*. Su virtud radica en que nos permite conocer la probabilidad asociada a cada uno de los diferentes valores \bar{Y} que es posible obtener en muestras de tamaño n cuando: 1) suponemos que el verdadero valor de la media poblacional es μ ; y 2) estimamos la desviación típica poblacional σ mediante la desviación típica muestral S_{n-1} .

Para que el estadístico T se ajuste apropiadamente al modelo de distribución de probabilidad t de *Student*, es necesario que la población muestreada sea *normal*. No obstante, con tamaños muestrales grandes, el ajuste de T a la distribución t de *Student* es lo suficientemente bueno incluso con poblaciones originales sensiblemente alejadas de la normalidad.

Para contrastar hipótesis sobre una media:

- ▶ Seleccionar la opción **Comparar medias > Prueba T para una muestra...** del menú **Analizar** para acceder al cuadro de diálogo *Prueba T para una muestra* que recoge la figura 13.3.

Figura 13.3. Cuadro de diálogo *Prueba T para una muestra*.



La lista de variables contiene un listado con todas las variables del archivo de datos excepto las que tienen formato de *cadena*. Para llevar a cabo un contraste con las especificaciones que el procedimiento tiene establecidas por defecto:

- ▶ Trasladar a la lista **Contrastar variables** la variable (o variables) cuya media poblacional se desea contrastar.
- ▶ Indicar en **Valor de prueba** el valor poblacional concreto que se desea contrastar. Este valor se utiliza para todas las variables seleccionadas en la lista **Contrastar variables**.
- ▶ Pulsar el botón **Aceptar**.

Cada variable seleccionada genera una *prueba T* acompañada de su correspondiente *nivel crítico bilateral* (el unilateral se obtiene dividiendo el bilateral por 2). El nivel crítico indica la probabilidad de obtener una media \bar{Y} tan alejada de μ como la de hecho obtenida (μ es el valor

propuesto en el cuadro de texto **Valor de prueba**). Si esa probabilidad es muy pequeña (generalmente, menor que 0,05), podemos rechazar la hipótesis nula de que la media poblacional es el valor propuesto.

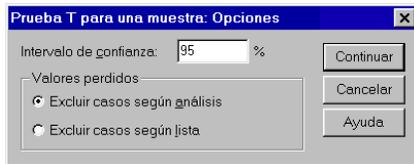
Los resultados que ofrece el *Visor* también incluyen el intervalo de confianza para la diferencia entre la media muestral \bar{Y} y el **Valor de prueba** (calculado al 95 por ciento). Los límites de confianza de este intervalo se calculan sumando y restando a la diferencia entre \bar{Y} y el **Valor de prueba** una cantidad que se obtiene multiplicando el error típico de la media ($\hat{\sigma}_{\bar{Y}}$) por el cuantil 97,5 de la distribución *t* de *Student* con $n-1$ grados de libertad (es decir, multiplicando el error típico de la media por el cuantil $100(1-\alpha/2) = 100(1-0,05/2) = 97,5$, donde α se refiere al nivel de significación).

Opciones

Las opciones del procedimiento **Prueba T para una muestra** permiten controlar algunos aspectos del análisis. Para modificar las opciones por defecto:

- ▶ Pulsar el botón **Opciones...** del cuadro de diálogo *Prueba T para una muestra* (ver figura 13.3) para acceder al subcuadro de diálogo *Prueba T para una muestra: Opciones* que ofrece la figura 13.4.

Figura 13.4. Subcuadro de diálogo *Prueba T para una muestra: Opciones*.



Intervalo de confianza: k %. Esta opción permite establecer, en escala porcentual, el *nivel de confianza* ($1-\alpha$) con el que deseamos obtener el intervalo de confianza para la diferencia entre la media muestral y el **Valor de prueba**. El valor de k es, por defecto, 95, pero es posible seleccionar cualquier otro valor comprendido entre 0,01 y 99,99.

Un intervalo de confianza sirve para tomar una decisión sobre la misma hipótesis nula que permite contrastar el estadístico T : cuando el nivel crítico bilateral asociado al estadístico T es menor que 0,05, el intervalo de confianza obtenido al 95 % no incluye el valor cero.

Valores perdidos. Podemos optar entre dos formas diferentes de tratar los casos con valores perdidos:

- **Excluir casos según análisis.** Esta opción excluye de cada análisis (de cada prueba T) los casos con valor perdido en la variable concreta que se está contrastando.
- **Excluir casos según lista.** Esta opción excluye de todos los análisis los casos con algún valor perdido en una cualquiera de las variables seleccionadas en la lista **Contrastar variables**.

Ejemplo (Comparar medias > Prueba T para una muestra)

Este ejemplo muestra cómo contrastar la hipótesis de que la media poblacional de la variable *ci* (cociente intelectual) vale 100. La variable *ci* la hemos generado con la función *RV.NORMAL* del procedimiento **Calcular**, utilizando una media de 100 y una desviación típica de 15 (ver apéndice).

- ▶ En el cuadro de diálogo *Prueba T para una muestra* (ver figura 13.3), seleccionar la variable *ci* y trasladarla a la lista **Contrastar variables**.
- ▶ Introducir el valor 100 en el cuadro de texto **Valor de prueba**.

Aceptando estos valores, el *Visor de resultados* ofrece la información que muestran las tablas 13.2.a y 13.2.b.

Tabla 13.2.a. Tabla de estadísticos del procedimiento *Prueba T para una muestra*.

	N	Media	Desviación típica	Error típ. de la media
Cociente intelectual	474	100,4630	14,5227	,6671

La tabla 13.2.a incluye el número de casos válidos sobre el que se basan los cálculos (474), la media de la variable *ci* (100,463), la desviación típica insesgada (14,5227) y el error típico de la media (0,6671; este error típico es el denominador de la *prueba T* y se obtiene dividiendo la desviación típica insesgada por la raíz cuadrada del número de casos).

Tabla 13.2.b. Tabla resumen del procedimiento *Prueba T para una muestra*.

	Valor de prueba = 100				
	t	gl	Sig. (bilateral)	Diferencia de medias	Intervalo de confianza para la diferencia
					Inferior
Cociente intelectual	,694	473	,488	,4630	-,8478 1,7737

La tabla 13.2.b ofrece un resumen de la prueba *T* encabezado con una línea que nos recuerda cuál es el valor propuesto para la media poblacional (*Valor de prueba* = 100). Las primeras columnas contienen el valor del estadístico ($t = 0,694$), sus grados de libertad ($gl = 473$) y el nivel crítico bilateral (*Significación bilateral* = 0,488). El nivel crítico muestra el grado de

compatibilidad entre el valor poblacional propuesto y la información muestral disponible: si el nivel crítico es pequeño (generalmente menor que 0,05), concluiremos que los datos se muestran incompatibles con la hipótesis de que el verdadero valor de la media poblacional es el propuesto. En nuestro ejemplo, el nivel crítico vale 0,488; puesto que es mayor que 0,05, no podemos rechazar como verdadero el valor propuesto. Podemos concluir, por tanto, que los datos muestrales pueden haber sido extraídos de una población con media 100.

Las siguiente columna de la tabla 13.2.b contiene la diferencia entre la media muestral y el valor de prueba (*Diferencia de medias* = 0,463). Esta diferencia es el numerador de la *prueba T*. Y a continuación aparecen los límites inferior (-0,8478) y superior (1,7737) del intervalo de confianza (calculado al 95 por ciento) para la diferencia entre la media muestral y el valor de prueba. Estos límites también permiten decidir sobre el valor propuesto para la media poblacional: si los límites incluyen el valor cero (como ocurre en nuestro ejemplo), podemos concluir que los datos muestrales son compatibles con el valor poblacional propuesto y, en consecuencia, mantener H_0 ; si los límites no incluyen el valor cero, debemos concluir que los datos son incompatibles con el valor propuesto y, consecuentemente, rechazar H_0 .

No hemos dicho nada sobre el supuesto de normalidad en que se basa la prueba *T*, pero sabemos que el cumplimiento de este supuesto sólo es exigible con muestras pequeñas. Con una muestra de 474 casos, como en nuestro ejemplo, el supuesto de normalidad carece de relevancia. No obstante, el SPSS permite contrastar la hipótesis de normalidad mediante el procedimiento **Explorar** ya estudiado en el capítulo 11.

Prueba T para muestras independientes

La prueba T para dos muestras independientes permite contrastar hipótesis referidas a la diferencia entre dos medias independientes. Tenemos dos poblaciones normales, con medias μ_1 y μ_2 , de cada una de las cuales seleccionamos una muestra aleatoria (de tamaños n_1 y n_2). Tras esto, utilizamos las medias muestrales \bar{Y}_1 e \bar{Y}_2 para contrastar la hipótesis de que las medias poblacionales μ_1 y μ_2 son iguales.

La prueba T que permite contrastar esta hipótesis de igualdad de medias no es otra cosa que una *tipificación de la diferencia entre las dos medias muestrales* que se obtiene restando a esa diferencia su valor esperado en la población y dividiendo el resultado por el error típico de la diferencia:

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}_{\bar{Y}_1 - \bar{Y}_2}}$$

El estadístico o prueba T tiene dos versiones que se diferencian en la forma concreta de estimar el error típico $\hat{\sigma}_{\bar{Y}_1 - \bar{Y}_2}$. Si suponemos que las varianzas poblacionales σ_1^2 y σ_2^2 son *iguales* (es decir, si suponemos $\sigma_1^2 = \sigma_2^2 = \sigma^2$), podemos estimar esa única varianza poblacional σ^2 utilizando una combinación ponderada de las varianzas (insesgadas) muestrales $S_{n_1-1}^2$ y $S_{n_2-1}^2$:

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_{n_1-1}^2 + (n_2 - 1)S_{n_2-1}^2}{n_1 + n_2 - 2}$$

Con esta estimación *combinada* de la varianza poblacional podemos obtener el error típico de la diferencia como:

$$\hat{\sigma}_{\bar{Y}_1 - \bar{Y}_2} = \hat{\sigma} \sqrt{1/n_1 + 1/n_2}$$

Procediendo de esta manera, el estadístico T resultante se distribuye según el modelo de probabilidad t de *Student* con $n_1 + n_2 - 2$ grados de libertad.

Si no podemos suponer que las varianzas poblacionales son *iguales*, entonces σ_1^2 debe estimarse mediante $S_{n_1-1}^2$ y σ_2^2 mediante $S_{n_2-1}^2$, en cuyo caso, el error típico de la diferencia podemos estimarlo mediante:

$$\hat{\sigma}_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{S_{n_1-1}^2/n_1 + S_{n_2-1}^2/n_2}$$

Al utilizar este error típico, el estadístico T sigue siendo una variable distribuida según el modelo de probabilidad t de *Student*, pero los grados de libertad de la distribución cambian. Necesitan ser estimados, para lo cual se utiliza una ecuación propuesta por Welch (1938):

$$gl = \frac{\left(S_{n_1-1}^2/n_1 + S_{n_2-1}^2/n_2 \right)^2}{\left(S_{n_1-1}^2/n_1 \right)^2 / (n_1 - 1) + \left(S_{n_2-1}^2/n_2 \right)^2 / (n_2 - 1)}$$

Para decidir si podemos o no suponer varianzas poblacionales iguales, el procedimiento *Prueba T para muestras independientes*, además de ofrecer las dos versiones del estadístico T , proporciona la prueba de *Levene* sobre igualdad de varianzas (esta prueba ya ha sido descrita en el capítulo 11, apartado *Cómo contrastar supuestos: homogeneidad de varianzas*). Supondremos o no varianzas iguales (y, por tanto utilizaremos una u otra versión del estadístico T) dependiendo de la conclusión a la que nos lleve la prueba de *Levene*.

Para comparar dos medias independientes:

- ▶ Seleccionar la opción **Comparar medias > Prueba T para dos muestras independientes...** del menú **Analizar** para acceder al cuadro de diálogo *Prueba T para dos muestras independientes* que recoge la figura 13.5.

Figura 13.5. Cuadro de diálogo *Prueba T para muestras independientes*.



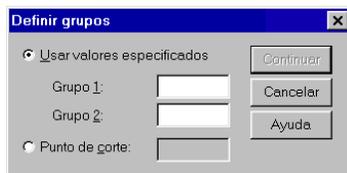
La lista de variables contiene un listado con las variables numéricas y de cadena corta del archivo de datos. Para llevar a cabo un contraste con las especificaciones que el procedimiento tiene establecidas por defecto:

- ▶ Trasladar a la lista **Contrastar variables** la variable (o variables) en la que se desea comparar los grupos (en el ejemplo de la figura 13.5 hemos seleccionado la variable *salini: salario inicial*). Todas estas variables deben tener formato numérico. Cada variable seleccionada genera una prueba *T* acompañada de su nivel crítico y del intervalo de confianza para la diferencia entre las medias.
- ▶ Trasladar al cuadro **Variable de agrupación** la variable que define los grupos que se desea comparar. Esta variable puede tener formato numérico o de cadena corta. (En el ejemplo de la figura 13.5 hemos seleccionado la variable *genero*, la cual define dos grupos: 1 = *hombres* y 2 = *mujeres*; la variable *genero* es una variable numérica obtenida a partir de la variable de cadena corta *sexo*; ver apéndice).

Definir grupos. Tras seleccionar una variable de agrupación, es necesario indicar cuáles son los códigos que definen los dos grupos que se desea comparar. Para ello:

- ▶ Pulsar el botón **Definir grupos...** para acceder al subcuadro de diálogo *Definir grupos* que muestra la figura 13.6.

Figura 13.6. Subcuadro de diálogo *Definir grupos*.



- **Usar valores especificados.** Si la variable de agrupación posee códigos que definen los dos grupos que deseamos comparar, esos códigos deben introducirse en los cuadros de texto **Grupo 1** y **Grupo 2**. Los casos que posean otros códigos serán excluidos del análisis. En el ejemplo de la figura 13.6 hemos introducido los códigos '1' y '2' (que son los códigos que, en el archivo *Datos de empleados*, corresponden a la variable *género*).

- **Punto de corte.** Si se desea utilizar como variable de agrupación una variable cuantitativa *continua*, esta opción permite introducir un valor como punto de corte: los casos con puntuación igual o mayor que el punto de corte forman un grupo; el resto de los casos forman el otro grupo. Esta opción no está disponible si, como variable de agrupación, se elige una variable con formato de cadena corta.

Opciones

Las opciones del cuadro de diálogo *Prueba T para muestras independientes* permiten controlar aspectos del análisis como el nivel de confianza con el que se desea trabajar y el tratamiento que se desea dar a los valores perdidos. Para modificar las opciones por defecto:

- ▶ Pulsar el botón **Opciones...** (ver figura 13.5) para acceder al subcuadro de diálogo *Prueba T para muestras independientes: Opciones* que recoge la figura 13.7.

Figura 13.7. Subcuadro de diálogo *Prueba T para muestras independientes: Opciones*.



Intervalo de confianza: k %. Esta opción permite establecer, en escala porcentual, el *nivel de confianza* ($1-\alpha$) con el que se desea obtener el intervalo de confianza para la diferencia de medias. El valor de k es, por defecto, 95, pero es posible seleccionar cualquier otro valor comprendido entre 0,01 y 99,99.

Valores perdidos. Podemos optar entre dos formas diferentes de tratar los casos con valores perdidos:

- **Excluir casos según análisis.** Esta opción excluye de cada análisis (de cada prueba T) los casos con valor perdido en la variable de agrupación o en la variable que se está contrastando en ese análisis.
- **Excluir casos según lista.** Esta opción excluye de todos los análisis (de todas las pruebas T solicitadas) los casos con algún valor perdido en la variable de agrupación o en una cualquiera de las variables incluidas en la lista **Contrastar variables**.

Ejemplo (Comparar medias > Prueba T para muestras independientes)

Este ejemplo muestra cómo contrastar hipótesis sobre dos medias poblacionales utilizando el procedimiento **Prueba T para muestras independientes**. En concreto, vamos a comparar las medias de *salini* (salario inicial) de los dos grupos definidos por la variable *sexo* ($h = \text{hombres}$ y $m = \text{mujeres}$).

- ▶ En el cuadro de diálogo *Prueba T para muestras independientes* (ver figura 13.5), trasladar la variable *salini* a la lista **Contrastar variables** y la variable *sexo* al cuadro **Variable de agrupación** (ver figura 13.5).
- ▶ Pulsar el botón **Definir grupos...** e introducir los códigos h y m en los cuadros de texto **Grupo 1** y **Grupo 2** (ver figura 13.6).

Aceptando estas selecciones, el *Visor* ofrece los resultados que muestran las tablas 13.3.a y 13.3.b (hemos rotado las cabeceras de las columnas de tabla 13.3.b para ajustar el tamaño de la tabla a las dimensiones de la página).

Tabla 13.3.a. Estadísticos del procedimiento *Prueba T para muestras independientes*.

Salario inicial

Sexo	N	Media	Desviación típica	Error típico de la media
Hombre	258	\$20,301.40	\$9,111.78	\$567.27
Mujer	216	\$13,091.97	\$2,935.60	\$199.74

La tabla 13.3.a muestra los grupos que se están comparando (*hombres* y *mujeres*) y, para cada grupo, el número de casos, la media y la desviación típica del salario inicial, y el error típico del salario inicial medio.

Tabla 13.3.b Tabla resumen del procedimiento *Prueba T para muestras independientes*.

Salario inicial

	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
	F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error tip de la diferencia	Intervalo de confianza para la media	
								Inferior	Superior
Asumiendo varianzas iguales	105,969	,000	11,152	472	,000	\$7,209.43	\$646.45	\$5,939.16	\$8,479.70
No asumiendo varianzas iguales			11,987	318,818	,000	\$7,209.43	\$601.41	\$6,026.19	\$8,392.67

La tabla 13.3.b ofrece, en primer lugar, el contraste de Levene (F) sobre homogeneidad o igualdad de varianzas. El resultado de este contraste es el que nos permite decidir si podemos o no suponer que las varianzas poblacionales son iguales: si la probabilidad asociada al estadístico de Levene es mayor que 0,05, podremos suponer que las varianzas poblacionales son iguales; si la probabilidad asociada al estadístico de Levene es menor que 0,05, rechazaremos la hipótesis de igualdad de varianzas y supondremos que son distintas.

Las columnas siguientes contienen el estadístico t , sus grados de libertad (gl), el nivel crítico bilateral (*Significación bilateral*), la diferencia entre el salario medio de cada grupo, el error típico de esa diferencia, y los límites inferior y superior del intervalo de confianza al 95 por ciento. Toda esta información está calculada tanto para el caso de varianzas poblacionales iguales (línea encabezada *Asumiendo varianzas iguales*) como para el caso de varianzas poblacionales distintas (línea encabezada *No asumiendo varianzas iguales*).

En el ejemplo, la probabilidad asociada al estadístico de Levene (0,000) es menor que 0,05, por lo que debemos rechazar la hipótesis de igualdad de varianzas y, consecuentemente, utilizar la información de la fila encabezada *No asumiendo varianzas iguales*: el estadístico t toma el valor 11,987 y tiene asociado un nivel crítico bilateral de 0,000. Este valor es justamente el que nos informa sobre el grado de compatibilidad existente entre la diferencia observada entre las medias muestrales de los grupos comprados y la hipótesis nula de que las medias

poblacionales son iguales. Puesto que $0,000$ es menor que $0,05$, podemos rechazar la hipótesis de igualdad de medias y, consecuentemente, concluir que el salario medio de hombres y mujeres no es el mismo.

Los límites del intervalo de confianza nos permiten estimar que la verdadera diferencia entre el salario medio de la población de hombres y el salario medio de la población de mujeres se encuentra entre $6.026,19$ y $8.392,67$ dólares. El hecho de que el intervalo obtenido no incluya el valor cero también nos permite rechazar la hipótesis de igualdad de medias.

Prueba T para muestras relacionadas

La prueba T para dos muestras relacionadas permite contrastar hipótesis referidas a la diferencia entre dos medias relacionadas.

Ahora disponemos de una población de diferencias con media μ_D , obtenida al restar las puntuaciones del mismo grupo de casos en dos variables diferentes o en la misma variable medida en dos momentos diferentes (de ahí que hablemos de muestras relacionadas). De esa población de diferencias extraemos una muestra aleatoria de tamaño n y utilizamos la media \bar{Y}_D de esas n diferencias para contrastar la hipótesis de que la media μ_D de la población de diferencias vale cero.

Desde el punto de vista estadístico, este contraste es idéntico al presentado en el apartado *Prueba T para una muestra*. La única diferencia existente entre ambos contrastes es que allí teníamos una muestra de puntuaciones obtenida al medir una sola variable y ahora tenemos dos muestras relacionadas (o una muestra de *pares* de puntuaciones) que convertimos en una sola muestra de *diferencias* restando las puntuaciones de cada par.

El estadístico o prueba T sigue siendo una tipificación de la media muestral de las diferencias \bar{Y}_D :

$$T = \frac{\bar{Y}_D - \mu_D}{\hat{\sigma}_{\bar{Y}_D}} = \frac{\bar{Y}_D - \mu_D}{S_D / \sqrt{n}}$$

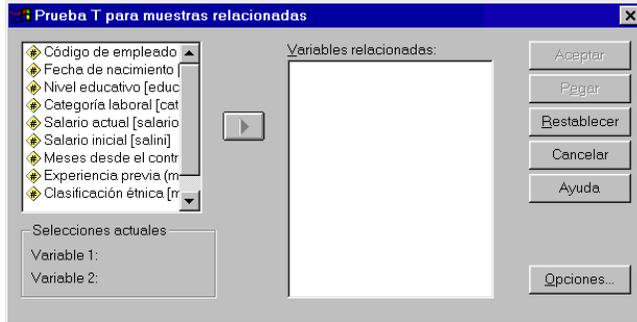
(S_D se refiere a la desviación típica insesgada de las n diferencias). Este estadístico T se distribuye según el modelo t de *Student* con $n-1$ grados de libertad y, por tanto, nos permite conocer la probabilidad asociada a los diferentes valores \bar{Y}_D que es posible obtener en muestras aleatorias de tamaño n .

Al igual que antes, para que el valor T se ajuste apropiadamente al modelo de distribución de probabilidad t de *Student*, es necesario que la población de diferencias sea *normal*. No obstante, con tamaños muestrales grandes el ajuste del estadístico T a la distribución t de *Student* es lo suficientemente bueno incluso con poblaciones originales alejadas de la normalidad.

Para contrastar hipótesis sobre dos medias relacionadas:

- ▶ Seleccionar la opción **Comparar medias > Prueba T para muestras relacionadas...** del menú **Analizar** para acceder al cuadro de diálogo *Prueba T para muestras relacionadas* que recoge la figura 13.8.

Figura 13.8. Cuadro de diálogo *Prueba T para muestras relacionadas*.



Lógicamente, la lista de variables sólo contiene variables con formato numérico. Para llevar a cabo un contraste sobre dos medias relacionadas con las especificaciones que el programa tiene establecidas por defecto:

- ▶ Trasladar a la lista **Variables relacionadas** las variables cuyas medias se desea comparar.

Las variables deben trasladarse a esta lista *por pares*. Es decir, es necesario marcar dos variables de la lista de variables para que el botón flecha esté activo. Pueden trasladarse tantos pares de variables como se desee.

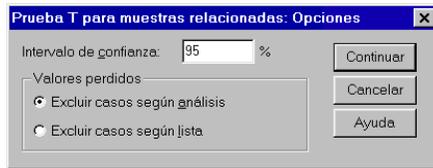
Una vez seleccionado uno o más pares de variables, el botón aceptar genera tantas pruebas *T* como pares seleccionados (incluyendo los correspondientes niveles críticos bilaterales y los intervalos de confianza para las diferencias).

Opciones

Las opciones del cuadro de diálogo *Prueba T para muestras relacionadas* permiten controlar el nivel de confianza con el que se desea trabajar y el tratamiento que se desea dar a los casos con valores perdidos. Para modificar las opciones:

- ▶ Pulsar el botón **Opciones...** del cuadro de diálogo *Prueba T para dos muestras relacionadas* (ver figura 13.8) para acceder al subcuadro de diálogo *Prueba T para muestras relacionadas: Opciones* que muestra la figura 13.9.

Figura 13.9. Cuadro de diálogo *Prueba T para muestras relacionadas: Opciones*.



Intervalo de confianza: k %. Esta opción permite establecer, en escala porcentual, el *nivel de confianza* ($1-\alpha$) con el que se desea obtener el intervalo de confianza para la diferencia de medias. El valor de k es, por defecto, 95, pero es posible introducir cualquier otro valor comprendido entre 0,01 y 99,99.

Valores perdidos. Podemos optar entre dos formas diferentes de tratar los casos con valores perdidos:

- **Excluir casos según análisis.** Esta opción excluye de cada análisis (de cada prueba T) los casos con valor perdido en cualquiera de las dos variables que están siendo contrastadas.
- **Excluir casos según lista.** Esta opción excluye de todos los análisis (de todas las pruebas T solicitadas) los casos con algún valor perdido en cualquiera de las variables seleccionadas en la lista **Variables relacionadas**.

Ejemplo (Comparar medias > Prueba T para muestras relacionadas)

Este ejemplo muestra cómo contrastar hipótesis sobre dos medias poblacionales utilizando el procedimiento **Prueba T para muestras relacionadas**. En concreto, vamos a comparar las medias de las variables *salini* (salario inicial) y *salario* (salario actual). Para ello:

- ▶ En el cuadro de diálogo *Prueba T para muestras relacionadas* (ver figura 13.8), seleccionar las variables *salini* y *salario* y trasladarlas a la lista **Variables relacionadas**.

Pulsando el botón **Aceptar**, el *Visor* ofrece los resultados que muestran las tablas 13.4.a, 13.4.b y 13.4.c.

Tabla 13.4.a. Estadísticos descriptivos del procedimiento *Prueba T para muestras relacionadas*.

		Media	N	Desviación típica	Error típ. de la media
Par 1	Salario actual	\$34,419.57	474	\$17,075.66	\$784.31
	Salario inicial	\$17,016.09	474	\$7,870.64	\$361.51

Tabla 13.4.b. Coeficiente de correlación de *Pearson*.

	N	Correlación	Sig.
Par 1 Salario actual y Salario inicial	474	0,88	.000

Tabla 13.4.c. Tabla resumen del procedimiento *Prueba T para muestras relacionadas*.

Par 1 Salario actual - Salario inicial

Diferencias relacionadas					t	gl	Sig. (bilateral)
Media	Desviación típica	Error típico de la media	Intervalo de confianza para la diferencia				
			Inferior	Superior			
\$17,403.48	\$10,814.62	\$496.73	\$16,427.41	\$18,379.56	35,036	473	,000

La tabla 13.4.a recoge, para cada variable, la media, el numero de casos, la desviación típica y el error típico de la media.

La tabla 13.4.b ofrece el coeficiente de correlación de Pearson entre ambas variables junto con el nivel crítico bilateral que le corresponde bajo la hipótesis de independencia (el coeficiente de correlación de Pearson será tratado en el capítulo 16).

La tabla 13.4.c incluye, en la primera mitad, tres estadísticos referidos a las *diferencias* entre cada par de puntuaciones: la media, la desviación típica y el error típico de la media. La siguiente columna contiene el intervalo de confianza para la diferencia entre las medias: podemos estimar, con una confianza del 95 por ciento, que la verdadera diferencia entre las medias de *salario* y *salini* se encuentra entre 16.427,41 y 18.379,56 dólares. La segunda mitad de la tabla informa sobre el valor del estadístico *t*, sus grados de libertad (*gl*) y el nivel crítico bilateral (*Sig. bilateral*). El valor del nivel crítico es muy pequeño (0,000), por lo que podemos rechazar la hipótesis de igualdad de medias y concluir que el *salario actual* medio es significativamente mayor que el *salario inicial* medio.

< Fin del capítulo 13 >