Capítulo 19
Análisis no paramétrico
El procedimiento *Pruebas no paramétricas*

En los capítulos 13 al 18 hemos estudiado una serie de procedimientos estadísticos diseñados para analizar variables cuantitativas: la prueba *T* para contrastar hipótesis sobre medias o coeficientes de regresión, el estadístico *F* del análisis de varianza y de la prueba de Levene, etc. Todos ellos coinciden en una serie de características:

1. Permiten contrastar hipótesis referidas a algún parámetro (μ, σ^2, ρ, β, etc.).
2. Exigen el cumplimiento de determinados supuestos sobre las poblaciones originales de las que se extraen los datos (generalmente normalidad y homocedasticidad).
3. Analizan datos obtenidos con una escala de medida de intervalo o razón.

Estas tres características combinadas permiten agrupar estos procedimientos estadísticos en una gran familia de técnicas de análisis denominada *contrastes paramétricos* (o *pruebas paramétricas*, en terminología afín a la del SPSS). Son, sin duda, las técnicas estadísticas más frecuentemente utilizadas por analistas e investigadores en todo tipo áreas científicas, pero su utilidad se ve reducida, fundamentalmente, por dos razones: por un lado, exigen el cumplimiento de algunos supuestos que en ocasiones pueden resultar demasiado exigentes; por otro, obligan a trabajar con unos niveles de medida que, especialmente en las ciencias sociales y de la salud, no siempre resulta fácil alcanzar.

Afortunadamente, los contrastes paramétricos no son los únicos disponibles. Existen contrastes que permiten poner a prueba hipótesis no referidas a parámetros poblacionales; existen también contrastes que no necesitan establecer supuestos exigentes sobre las poblaciones de donde se extraen las muestras; y existen, por último, contrastes que no necesitan trabajar con datos obtenidos con una escala de medida de intervalo o razón. Esta otra familia de contrastes se conoce con el nombre de *contrastes no paramétricos* (o *pruebas no paramétricas*).

Algunos autores utilizan el término *no paramétricos* para referirse únicamente a los contrastes que no plantean hipótesis sobre parámetros y que se limitan a analizar las propiedades nominales u ordinales de los datos, y añaden el término *de distribución libre* para referirse a los contrastes que no necesitan establecer supuestos (o establecen supuestos poco exigentes, como simetría o continuidad) sobre las poblaciones originales de las que se extraen las muestras. Pero lo cierto es que el incumplimiento de cualquiera de las tres características señaladas al principio puede ser considerada suficiente para caracterizar a un contraste como *no paramétrico*. De esta forma, podemos 1) utilizar la denominación genérica de *no paramétricos* para todos aquellos contrastes que no se ajustan a una cualquiera de las tres características de los contrastes *paramétricos* y, por tanto, 2) englobar en ese término genérico a los contrastes *de distribución libre*.

Más allá del acuerdo que pueda existir sobre esta cuestión, poner el énfasis en el *nivel de medida* de los datos contribuye a simplificar notablemente la clasificación e identificación de las distintas técnicas de análisis de datos. Por tanto, podemos 1) clasificar los contrastes de acuerdo con el *tipo de datos* que permiten analizar (independientemente del tipo de hipótesis que permitan contrastar e independientemente de los supuestos que sea necesario establecer) y 2) llamarlos, a todos ellos, *no paramétricos* siempre que no se ajusten a una cualquiera de las tres características de los contrastes *paramétricos*.

Este capítulo ofrece una descripción de las técnicas de análisis que el SPSS clasifica como *pruebas no paramétricas*. Todas ellas pueden considerarse *no paramétricas* utilizando el criterio de que no plantean hipótesis sobre parámetros, o el de que analizan datos obtenidos con una escala de medida débil (o mejor, datos que, aun estando medidos con una escala de intervalo o razón, se analizan aprovechando sólo sus propiedades nominales u ordinales); y muchas de ellas pueden considerarse de *distribución libre* utilizando el criterio de que no establecen supuestos demasiado exigentes sobre las poblaciones originales de donde se muestrea.

Todas estas pruebas se encuentran en la opción **Pruebas no paramétricas** del menú **Analizar**. Y aparecen ordenadas por el número de muestras que permiten analizar (el módulo *Pruebas exactas* incluye dos pruebas adicionales no incluidas en el módulo *Base*: la prueba de *Jonckheere-Terpstra* y la prueba de *homogeneidad marginal*):

- ***Pruebas para una muestra***: *Chi-cuadrado* (bondad de ajuste con variables categóricas), *Binomial* (proporciones y cuantiles), *Rachas* (aleatoriedad) y *Kolmogorov-Smirnov* (bondad de ajuste con variables cuantitativas).
- ***Pruebas para dos muestras independientes***: *U de Mann-Whitney*, *Kolmogorov-Smirnov*, *Reacciones extremas de Moses* y *Rachas de Wald-Wolfowitz*.
- ***Pruebas para varias muestras independientes***: *H de Kruskal-Wallis* y *Mediana*.
- ***Pruebas para dos muestras relacionadas***: *Wilcoxon*, *Signos* y *McNemar*.
- ***Pruebas para varias muestras relacionadas***: *Friedman*, *W de Kendall* y *Q de Cochran*.

Pruebas para una muestra

Prueba *Chi-cuadrado* para una muestra

La prueba *chi-cuadrado* para una muestra permite averiguar si la distribución empírica de una variable categórica se ajusta o no (se parece o no) a una determinada distribución teórica (uniforme, binomial, multinomial, etc.). Esta hipótesis de ajuste, o mejor, de bondad de ajuste, se pone a prueba utilizando un estadístico originalmente propuesto por Pearson (1900; ver también Cochran, 1952) para comparar las frecuencias observadas o empíricas con las esperadas o teóricas de cada categoría, es decir, un estadístico diseñado para comparar las frecuencias de hecho obtenidas en una muestra concreta (frecuencias observadas: n_i) con las frecuencias que deberíamos encontrar si la variable realmente siguiera la distribución teórica propuesta en la hipótesis nula (frecuencias esperadas: m_i):

$$X^2 = \sum_i \frac{(n_i - m_i)^2}{m_i}$$

La frecuencias esperadas m_i se obtienen multiplicando la probabilidad teórica de cada categoría π_i (la que corresponde a cada categoría de acuerdo con la hipótesis nula) por el número de casos válidos: $n\pi_i$. Si no existen casillas vacías y el número de frecuencias esperadas menores de 5 no superan el 20 % del total de frecuencias esperadas (Cochran, 1952), el estadístico X^2 se distribuye según el modelo de probabilidad *chi-cuadrado* con k–1 grados de libertad (donde k se refiere al número de categorías de la variable cuyo ajuste se está intentando evaluar).

Para obtener la prueba *chi-cuadrado*:

- Seleccionar la opción **Pruebas no paramétricas...** > **Chi-cuadrado** del menú **Analizar** para acceder al cuadro de diálogo *Prueba de chi-cuadrado* que muestra la figura 19.1.

Figura 19.1. Cuadro de diálogo *Prueba Chi-cuadrado*.

La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para contrastar la hipótesis de bondad de ajuste referida a una variable categórica:

- Trasladar esa variable a la lista **Contrastar variables**. Si se selecciona más de una variable, el SPSS ofrece tantos contrastes como variables.

Rango esperado. Es posible decidir qué rango de valores de la variable seleccionada deben tenerse en cuenta en el análisis:

- **Obtener de los datos**. Cada valor distinto de la variable se considera una categoría para el análisis.
- **Usar rango especificado**. Sólo se tienen en cuenta los valores comprendidos entre los límites especificados en las cajas **Inferior** y **Superior**. Los valores no incluidos en esos límites se excluyen.

Valores esperados. Las opciones de este recuadro sirven para hacer explícitas las frecuencias esperadas con las que se desea comparar las observadas:

- **Todas las categorías iguales**. Las frecuencias esperadas se obtienen dividiendo el número total de casos válidos entre el número de categorías de la variable. Equivale a efectuar el ajuste a una distribución uniforme.

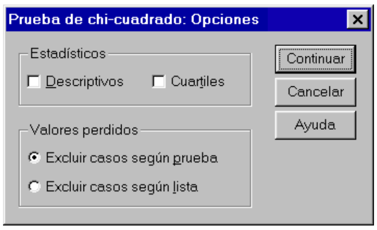
- **Valores**. Esta opción permite definir frecuencias esperadas concretas. Es importante tener en cuenta que los valores que se introducen se interpretan como proporciones, no como frecuencias absolutas. Deben introducirse tantos valores como categorías: el SPSS divide cada valor por la suma de todos los valores. Así, por ejemplo, si una variable tiene dos categorías y se introducen los enteros 6 y 4, el SPSS interpreta que la frecuencia esperada de la primera categoría es 6/10 del número de casos válidos y que la frecuencia esperada de la segunda categoría es 4/10 del número de casos válidos. De esta forma, resulta fácil definir, por ejemplo, las frecuencias esperadas correspondientes a una distribución binomial o multinomial.

  El orden en el que se introducen los valores es muy importante, pues la secuencia introducida se hace corresponder con las categorías de la variable cuando éstas se encuentran ordenadas de forma ascendente.

El botón **Opciones...** permite obtener algunos estadísticos descriptivos y decidir qué tratamiento se desea dar a los valores perdidos:

- Pulsar el botón **Opciones...** para acceder al cuadro de diálogo *Prueba de chi-cuadrado: Opciones* que muestra la figura 19.2.

Figura 19.2. Subcuadro de diálogo *Prueba Chi-cuadrado: Opciones*.



Estadísticos. Las opciones de este recuadro permiten obtener algunos estadísticos descriptivos:

- **Descriptivos**. Ofrece el número de casos válidos, la media, la desviación típica, el valor mínimo y el valor máximo.
- **Cuartiles**. Ofrece los centiles 25, 50 y 75.

Conviene señalar que estos estadísticos no siempre tendrán sentido, pues la prueba *chi-cuadrado* se utiliza generalmente con variables categóricas. Para contrastar la hipótesis de bondad de ajuste con variables cuantitativas es preferible utilizar la prueba de Kolmogorov-Smirnov.

Valores perdidos. Este recuadro permite decidir qué tratamiento se desea dar a los valores perdidos en el caso de que se haya seleccionado más de una variable:

- **Excluir casos según prueba**. Se excluyen de cada contraste los casos con valor perdido en la variable que se está contrastando. Es la opción por defecto.
- **Excluir casos según pareja**. Se excluyen de todos los contrastes solicitados los casos con algún valor perdido en cualquiera de las variables seleccionadas.

Ejemplo (pruebas no paramétricas > Chi-cuadrado)

Este ejemplo muestra cómo utilizar la prueba *chi-cuadrado* para contrastar la hipótesis de bondad de ajuste con una muestra (una variable). Para ello, vamos a utilizar la variable *educ* (nivel educativo) con intención de averiguar si las frecuencias de las categorías de esa variable se ajustan a una distribución uniforme:

- ▶ En el cuadro de diálogo *Prueba de chi-cuadrado* (ver figura 19.1), seleccionar la variable *educ* (nivel educativo) y trasladarla, mediante el botón flecha, a la lista **Contrastar variables**.
- ▶ Pulsar el botón **Opciones...** y marcar las opciones **Descriptivos** y **Cuartiles**.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran las tablas 19.1, 19.2 y 19.3.

La tabla 19.1 recoge la información descriptiva solicitada al marcar las opciones **Descriptivos** y **Cuartiles**: tenemos una muestra de 474 casos, con un promedio de años de formación académica de 13,49 y una desviación típica de 2,88, y con un rango de valores (años) que oscila entre 8 y 21. Los cuartiles indican, por ejemplo, que la mitad de los sujetos tiene al menos 12 años de estudios y que el 25 % de los sujetos tiene más de 15 años de estudios.

Tabla 19.1. Estadísticos descriptivos.

						Percentiles		
	N	Media	Desviación típica	Mínimo	Máximo	25	50 (Mediana)	75
Nivel educativo	474	13,49	2,88	8	21	12,00	12,00	15,00

La segunda tabla de resultados (tabla 19.2) contiene las frecuencias observadas y las esperadas, así como las diferencias entre ambas (*residual*).

Tabla 19.2. Frecuencias observadas y esperadas.

	N observado	N esperado	Residual
8	53	47,4	5,6
12	190	47,4	142,6
14	6	47,4	-41,4
15	116	47,4	68,6
16	59	47,4	11,6
17	11	47,4	-36,4
18	9	47,4	-38,4
19	27	47,4	-20,4
20	2	47,4	-45,4
21	1	47,4	-46,4
Total	474		

La tabla 19.3, por último, ofrece la información necesaria para tomar una decisión sobre la hipótesis de bondad de ajuste: el valor del estadístico *chi-cuadrado* (724,692), sus grados de libertad (*gl* = número de categorías menos uno) y su nivel crítico (*Sig.* = 0,000). Puesto que el nivel crítico es menor que 0,05, podemos rechazar la hipótesis de bondad de ajuste y concluir que la variable *educ* (nivel educativo) no se ajusta a una distribución uniforme.

Tabla 19.3. Estadístico *chi-cuadrado*.

	Nivel educativo
Chi-cuadrado	724,692
gl	9
Significación asintótica	,000

Prueba binomial

En la mayoría de las áreas de conocimiento es relativamente frecuente encontrarse con variables dicotómicas o dicotomizadas, es decir, con variables categóricas que sólo toman dos valores: éxito–fracaso, a favor–en contra, tratados–no tratados, recuperados–no recuperados, aprobados–suspensos, etc. Podemos llamar, de forma genérica, *acierto* y *error* a los dos niveles de una variable de este tipo. La prueba *binomial* permite averiguar si una variable dicotómica sigue o no un determinado modelo de probabilidad. En concreto, permite contrastar la hipótesis de que las proporción observada de *aciertos* se ajusta a la proporción teórica de una distribución binomial (lo cual se traduce, según veremos, en la posibilidad de contrastar hipótesis sobre proporciones y sobre cuantiles). John Arbuthnott (1710) fue el primero en utilizar este procedimiento para demostrar que la proporción de varones nacidos en Londres en un determinado periodo de tiempo era significativamente mayor que la proporción de mujeres.

Si extraemos muestras aleatorias de tamaño n y, en cada muestra, definimos la variable X = "número de aciertos en las n extracciones", tendremos una variable aleatoria distribuida, si la proporción de aciertos (π) permanece constante en cada extracción, según el modelo de probabilidad binomial, con parámetros n = "número de extracciones" y π = "proporción de aciertos". Podemos, por tanto, utilizar las probabilidades de la distribución binomial para conocer la probabilidad exacta asociada a cada uno de los valores de la variable X.

Ademas, a medida que n aumenta, la distribución de X se aproxima a la distribución normal con parámetros: $E(X) = n\pi$ y $\sigma_X = \sqrt{n\pi(1-\pi)}$. De modo que la variable[*]:

$$Z = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}}$$

se distribuirá según el modelo de probabilidad normal $N(0, 1)$. Podemos, también, por tanto, utilizar la distribución normal para conocer las probabilidades asociadas a los valores de X.

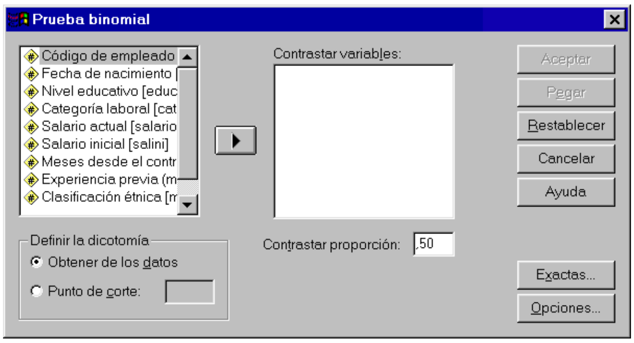
El SPSS utiliza ambas soluciones. Con muestras pequeñas ($n \leq 25$), utiliza la distribución binomial para obtener las probabilidades exactas asociadas a los valores del estadístico X. Con muestras grandes ($n > 25$) utiliza la distribución normal para obtener las probabilidades asociadas a los valores del estadístico Z (y, consecuentemente, las probabilidades aproximadas asociadas al estadístico X).

[*] El SPSS utiliza *la corrección por continuidad*, que consiste en sumar (si X es menor que $n\pi$) o restar (si X es mayor que $n\pi$) 0,5 puntos a X para hacer el contraste algo más conservador: $Z = [X \pm 0{,}5 - n\pi] \big/ \sqrt{n\pi(1-\pi)}$.

Para obtener la prueba *binomial*:

▶ Seleccionar la opción **Pruebas no paramétricas** > **Binomial...** del menú **Analizar** para acceder al cuadro de diálogo *Prueba binomial* que muestra la figura 19.3.

Figura 19.3. Cuadro de diálogo *Prueba binomial*.



La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para obtener la prueba *binomial*:

▶ Trasladar una o más variables a la lista **Contrastar variables**. Si se selecciona más de una variable, el SPSS ofrece un contraste por cada variable seleccionada.

Definir dicotomía. Las opciones de este recuadro permiten definir qué valores de la variable seleccionada van a utilizarse como categorías:

○ **Obtener de los datos**. Si la variable seleccionada es dicotómica, esta opción deja las cosas como están, es decir, deja que sean los propios valores de la variable los que definan la dicotomía. En ese caso, la hipótesis que permite contrastar la prueba *binomial* es si la proporción observada en la primera categoría se ajusta (se parece) a la proporción teórica propuesta en **Contrastar proporción**.

○ **Punto de corte**. Si la variable seleccionada no es dicotómica es necesario dicotomizarla. Para ello, debe indicarse el valor concreto que se utilizará para efectuar el corte: los valores menores o iguales que el punto de corte constituyen el primer grupo y los valores mayores el segundo.

Esta opción es extremadamente útil cuando lo que interesa es contrastar hipótesis sobre la mediana o sobre algún otro cuantil. Es decir, esta opción permite obtener los contrastes conocidos en la literatura estadística como *prueba de los signos* y *prueba de los cuantiles* (Ver San Martín y Pardo, 1989, págs 91-97). Si queremos contrastar, por ejemplo, la hipótesis de que la mediana del *salario inicial* es 25.000 dólares (*prueba de los signos*), podemos utilizar el valor 25.000 como **Punto de corte** y 0,5 (la proporción de casos acumulados hasta la mediana) como valor del contraste en **Contrastar proporción**. O si queremos contrastar la hipótesis de que el centil 80 del *salario inicial* vale 40.000 dólares (*prueba de los cuantiles*), podemos utilizar 40.000 como **Punto de corte** y 0,80 (la proporción de casos acumulados hasta el centil 80) como valor del contraste en la caja **Contrastar proporción**.

Así pues, las opciones del recuadro **Definir dicotomía** permiten decidir, entre otras cosas, qué tipo de contraste se desea llevar a cabo: sobre una *proporción* (si la variable es dicotómica) o sobre la *mediana* o cualquier otro *cuantil* (si la variable es al menos ordinal).

Contrastar proporción. Esta caja permite especificar el valor poblacional propuesto en la hipótesis nula. Por defecto, se asume que la variable dicotómica seleccionada sigue el modelo de distribución de probabilidad binomial con $\pi = 0{,}5$. Pero este valor de prueba puede cambiarse introduciendo un valor entre 0,001 y 0,999.

De las dos categorías de la variable dicotómica, la primera (aquella a la que corresponde el código menor) es la que se toma como categoría de referencia. Teniendo esto en cuenta:

- ***Si el valor de prueba es 0,5***, el SPSS interpreta que el contraste es *bilateral* y obtiene el nivel crítico multiplicando por dos la probabilidad de encontrar un número de casos igual o mayor que el de la categoría de referencia (si la proporción de casos de la categoría de referencia es mayor que 0,5) o multiplicando por dos la probabilidad de encontrar un número de casos igual o menor que el de la categoría de referencia (si la proporción de casos de la categoría de referencia es menor que 0,5).
- ***Si el valor de prueba es distinto de 0,5***, el SPSS interpreta que el contraste es *unilateral* y ofrece el nivel crítico resultante de calcular la probabilidad de encontrar un número de casos igual o mayor que el de la categoría de referencia (si la proporción de casos de la categoría de referencia es mayor que el valor de prueba; contraste unilateral derecho) o la probabilidad de encontrar un número de casos igual o menor que el de la categoría de referencia (si la proporción de casos de la categoría de referencia es menor que el valor de prueba; contraste unilateral izquierdo).

El botón **Opciones...** conduce a un subcuadro de diálogo idéntico al de la figura 19.2 que permite obtener algunos estadísticos descriptivos y decidir qué tratamiento se desea dar a los valores perdidos

Ejemplo (Pruebas no paramétricas > Binomial)

Este ejemplo muestra cómo utilizar la prueba *binomial* para contrastar la hipótesis de bondad de ajuste referida a una variable dicotómica. Para ello, vamos a utilizar la variable *minoría* (clasificación étnica). Si suponemos que el 70 % de los habitantes de EEUU es de raza blanca, puede resultar interesante averiguar si ese porcentaje se mantiene en la entidad bancaria a la que se refiere el archivo *Datos de empleados*. Para ello:

▶ En el cuadro de diálogo *Prueba binomial* (ver figura 19.3), seleccionar la variable *minoría* (clasificación étnica) y trasladarla, mediante el botón flecha, a la lista **Contrastar variables**.

▶ Introducir el valor 0,70 en la caja **Contrastar proporción** para especificar el valor de prueba.

▶ Puesto que la variable es dicotómica, dejar marcada la opción **Obtener de los datos** del recuadro **Definir dicotomía**.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestra la tabla 19.4. La tabla comienza identificando la variable que se esta utilizando en el contraste y los dos grupos que definen la dicotomía: *grupo 1 = minoría no* (blancos) y *grupo 2 = minoría sí* (no blancos).

Tabla 19.4. Prueba *binomial*.

		Categoría	N	Proporción observada	Proporción de prueba	Sig. asintót. (unilateral)
Clasificación étnica	Grupo 1	No	370	,780591	,7	,000[a]
	Grupo 2	Sí	104	,2		
	Total		474	1,0		

a. Basado en la aproximación Z.

Según sabemos, el SPSS toma como categoría de es la categoría con el código más pequeño. Por tanto, como los códigos de la variable *minoría* son 0 y 1, la categoría de referencia es la categoría con el código 0, es decir, la categoría *minoría = no*. La proporción de casos en esa categoría es 0,78 (=370/474) y la proporción de prueba es 0,7. Es muy importante fijarse en el tamaño de estas proporciones: puesto que el valor de prueba es distinto de 0,5 y la proporción de la categoría de referencia es mayor que el valor de prueba (0,78 > 0,7), el SPSS interpreta que el contraste es unilateral derecho y ofrece, como nivel crítico, la probabilidad de obtener, con $n = 474$ y $\pi = 0{,}7$, un número de casos igual o mayor que 370 (el número de casos de la categoría de referencia). Ese nivel crítico (*Significación asintótica unilateral*) vale 0,000, por lo que podemos rechazar la hipótesis nula de bondad de ajuste ($\pi \leq 0{,}7$) en favor de la alternativa ($\pi > 0{,}7$) y concluir que la verdadera proporción poblacional es mayor que 0,7.

Dado que el tamaño muestral es mayor que 25, la solución ofrecida se basa en la aproximación normal. Recordemos que sólo se utilizan las probabilidades exactas de la distribución binomial con tamaños muestrales no superiores a 25.

Prueba de las *rachas*

La prueba de las *rachas* sirve para determinar si una muestra de observaciones es o no aleatoria, es decir, si las observaciones de una determinada secuencia son independientes entre sí. En una serie temporal, por ejemplo, las observaciones no son aleatorias: lo que ocurre con una observación cualquiera depende, generalmente, de las características de alguna observación anterior. En una muestra aleatoria, por el contrario, debemos esperar que lo que ocurra con una observación cualquiera sea independiente de las características de la(s) observación(es) anterior(es).

El concepto de *racha* hace referencia a una secuencia de observaciones de un mismo tipo. Supongamos que lanzamos una moneda al aire 10 veces y que obtenemos el siguiente resultado: CCCXCCXXXC. Tendremos 5 rachas: CCC, X, CC, XXX y C. A simple vista, el resultado obtenido parece *aleatorio*. Pero si en lugar de ese resultado hubiéramos obtenido este otro: CCCCCXXXXX (sólo dos rachas) resultaría fácil ponernos de acuerdo en que la secuencia obtenida no parece aleatoria. Como tampoco parece aleatoria una secuencia con demasiadas rachas: CXCXCXCXCX (10 rachas).

Pues bien, la prueba de las rachas permite determinar si el número de rachas (R) observado en una determinada muestra de tamaño n es lo suficientemente grande o lo suficientemente pequeño como para poder rechazar la hipótesis de independencia (o aleatoriedad) entre las observaciones[*].

Para obtener el número de rachas es necesario que las observaciones estén clasificadas en dos grupos exhaustivos y mutuamente exclusivos (variable dicotómica). Si no lo están, deberemos utilizar algún criterio (mediana, media, moda, etc.) para hacer que lo estén.

Una vez clasificadas las n observaciones en dos grupos (de tamaños n_1 y n_2), el SPSS utiliza una tipificación[**] del número de rachas (R) para contrastar la hipótesis de aleatoriedad o independencia:

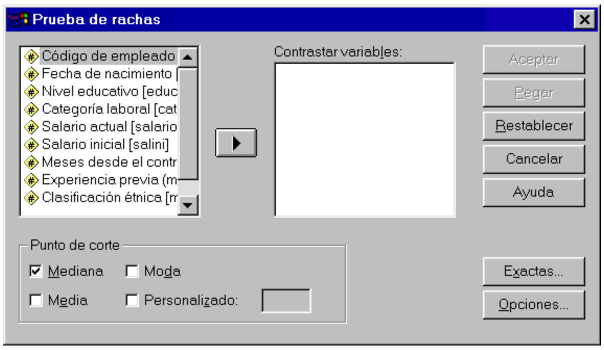
$$Z = \frac{R - E(R)}{\sigma_R}$$

donde: $E(R) = (2n_1n_2)/(n+1)$ y $\sigma_R = \sqrt{[2n_1n_2(2n_1n_2 - n)]/[n^2(n-1)]}$. El estadístico Z se distribuye según el modelo de probabilidad normal $N(0, 1)$. El SPSS ofrece el nivel crítico bilateral resultante de multiplicar por 2 la probabilidad de encontrar un número de rachas igual o menor que el encontrado (si $R < E(R)$), o un número de rachas igual o mayor que el encontrado (si $R > E(R)$).

Para obtener la prueba de las *rachas*:

▶ Seleccionar la opción **Pruebas no paramétricas** > **Rachas...** del menú **Analizar** para acceder al cuadro de diálogo *Prueba de las rachas* que muestra la figura 19.4.

Figura 19.4. Cuadro de diálogo *Prueba de las rachas.*



[*] Conviene no confundir la hipótesis de aleatoriedad con la hipótesis de bondad de ajuste estudiada a propósito de la prueba *binomial*. Obtener 5 caras y 5 cruces al lanzar una moneda 10 veces es un resultado que se ajusta perfectamente a la hipótesis de equiprobabilidad ($\pi_{cara} = \pi_{cruz} = 0{,}5$), pero si las 5 caras salen al principio y las cinco cruces al final, esto haría dudar de la hipótesis de independencia o aleatoriedad.

[**] Si el tamaño muestral es menor que 50, el estadístico Z se obtiene utilizando la *corrección por continuidad* de la siguiente manera:

- Si $R - E(R) < -0{,}5$, se suma 0,5 a R. Es decir: $Z = [R + 0{,}5 - E(R)]/\sigma_R$.
- Si $R - E(R) > 0{,}5$, se resta 0,5 a R. Es decir: $Z = [R - 0{,}5 - E(R)]/\sigma_R$.
- Si $|R - E(R)| \leq 0{,}5$, $Z = 0$.

La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para contrastar la hipótesis de aleatoriedad o independencia referida a una variable:

- ▶ Trasladar esa variable a la lista **Contrastar variables**. Si se selecciona más de una variable, el SPSS ofrece un contraste por cada variable seleccionada.

Punto de corte. Recordemos que para obtener el número de rachas es necesario que las observaciones estén clasificadas en dos grupos. Si no lo están, debemos utilizar algún criterio para hacer que lo estén. Si se desea contrastar la hipótesis de independencia referida a una *variable cuantitativa*, podemos utilizar como criterio de dicotomización (como punto de corte) la **Mediana**, la **Moda** o la **Media**. En ese caso, los valores más pequeños que el punto de corte pasan a formar parte del primer grupo y los valores iguales o mayores que el punto de corte pasan a formar parte del segundo grupo. Si se desea contrastar la hipótesis de independencia referida a una *variable categórica* puede utilizarse como punto de corte la opción **Personalizado**. Si la variable es, por ejemplo, dicotómica, con códigos 0 y 1, podemos utilizar como punto de corte el valor 0,5 (o cualquier otro comprendido entre 0 y 1, incluido el 1), de modo que los casos con código 0 pasen a formar parte del primer grupo y los casos con valor 1 pasen a formar parte del segundo grupo.

El botón **Opciones...** conduce a un subcuadro de diálogo idéntico al de la figura 19.2 que permite obtener algunos estadísticos descriptivos y decidir qué tratamiento se desea dar a los valores perdidos.

Ejemplo (Pruebas no paramétricas > Rachas)

Este ejemplo muestra cómo utilizar la prueba de las *rachas* para contrastar la hipótesis de independencia referida a la variable *salario inicial*.

- ▶ En el cuadro de diálogo *Prueba de las rachas* (ver figura 19.4), seleccionar la variable *salario inicial* y trasladarla, mediante el botón flecha, a la lista **Contrastar variables**.
- ▶ Dejar marcada la opción **Mediana** del recuadro **Punto de corte** para categorizar la variable utilizando como criterio la mediana.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestra la tabla 19.5. La tabla comienza indicando el punto de corte utilizado para la dicotomización: *Valor de prueba* = 15.000. Una llamada de nota a pie de tabla nos recuerda que ese punto de corte es la mediana.

Tabla 19.5. Prueba de las rachas.

	Salario inicial
Valor de prueba[a]	$15,000.00
Casos < Valor de prueba	212
Casos >= Valor de prueba	262
Casos en total	474
Número de rachas	150
Z	-7,939
Sig. asintót. (bilateral)	,000

a. Mediana

A continuación aparece el número de casos del primer grupo (*Casos < Valor de prueba* = 212), el número de casos del segundo grupo (*Casos >= Valor de prueba* = 262), el número de casos válidos de la muestra utilizada (*Casos en total* = 474) y el *número de rachas* computadas (150). La tabla ofrece, por último, el valor del estadístico de contraste (Z = –7,939) y su nivel crítico (*Significación asintótica bilateral* = 0,000). Puesto que el nivel crítico es muy pequeño (menor que 0,05), podemos rechazar la hipótesis de independencia y concluir que la secuencia de observaciones estudiada no es aleatoria.

Prueba de Kolmogorov-Smirnov para una muestra

Al igual que las pruebas *chi-cuadrado* para una muestra y *binomial*, la prueba de *Kolmogorov-Smirnov* (K-S) para una muestra (Kolmogorov, 1933) es una prueba de bondad de ajuste: sirve para contrastar la hipótesis nula de que la distribución de una variable se ajusta a una determinada distribución teórica de probabilidad. Pero a diferencia de las primeras, que han sido diseñadas más bien para evaluar el ajuste de variables categóricas, la prueba de K-S para una muestra se adapta mejor a situaciones en las interesa evaluar el ajuste de variables cuantitativas.

Para contrastar la hipótesis nula de bondad de ajuste, la prueba de K-S se basa en la comparación de dos funciones de distribución (o funciones de probabilidad acumuladas): una función de distribución empírica $F(X_i)$ y una función de distribución teórica $F_0(X_i)$.

Para obtener la función de distribución empírica $F(X_i)$ se comienza ordenando los valores de X_i de forma ascendente, es decir, desde el valor más pequeño $X_{[1]}$ hasta el más grande $X_{[n]}$. Tras esto, la función de distribución empírica para cada valor de X_i se obtiene de la siguiente manera: $F(X_i) = i / n$ (i se refiere al rango correspondiente a cada observación).

La forma de obtener la función de distribución teórica depende de la distribución concreta propuesta en la hipótesis. Si la distribución propuesta es, por ejemplo, la *uniforme*, la función de distribución teórica para cada valor de X_i se obtiene así: $F_0(X_i) = (X_i – X_{[1]})/(X_{[n]} – X_{[1]})$. Si la distribución teórica propuesta es, por ejemplo, la de Poisson, entonces la función de distribución teórica se obtiene así: $F_0(X_i) = \sum_{l=0}^{i}\left[e^{-\lambda}\lambda/l!\right]$. Etc.

Una vez obtenidas las distribuciones empírica y teórica, el estadístico de K-S se calcula a partir de la diferencia D_i más grande existente entre $F(X_i)$ y $F_0(X_i)$:

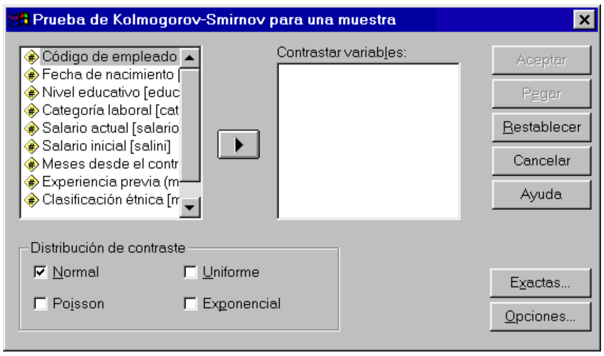
$$Z_{K-S} = máx\,|D_i|\,\sqrt{n}$$

Este estadístico Z se distribuye según el modelo de probabilidad normal $N(0, 1)$. El SPSS utiliza el método de Smirnov (1948) para obtener las probabilidades concretas asociadas a los valores del estadístico Z. Este método difiere del estándar (basado en las probabilidades de la curva normal estandarizada), pero es equivalente.

Para obtener la prueba de *Kolmogorov-Smirnov* para una muestra:

▶ Seleccionar la opción **Pruebas no paramétricas** > **K-S de una muestra...** del menú **Analizar** para acceder al cuadro de diálogo *Prueba de Kolmogorov-Smirnov para una muestra* que recoge la figura 19.5.

Figura 19.5. Cuadro de diálogo *Prueba de Kolmogorov-Smirnov para una muestra*.



La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para contrastar la hipótesis de bondad de ajuste referida a una variable:

- Trasladar esa variable a la lista **Contrastar variables**. Si se selecciona más de una variable, el SPSS ofrece un contraste por cada variable seleccionada.

Distribución de contraste. Las opciones de este apartado permiten elegir la distribución teórica a la cual se desea ajustar la distribución empírica de la variable seleccionada: **Normal**, **Uniforme**, **Poisson** y **Exponencial** (puede seleccionarse más de una). Los parámetros de las diferentes distribuciones se estiman a partir de los datos. No es posible obtener el ajuste a una distribución normal si la varianza de la variable vale cero. Ni a una distribución de Poisson si la media de la variable vale cero o los valores no son, todos ellos, enteros no negativos.

El botón **Opciones...** conduce a un subcuadro de diálogo idéntico al de la figura 19.2 que permite obtener algunos estadísticos descriptivos y decidir qué tratamiento se desea dar a los valores perdidos.

Ejemplo (Pruebas no paramétricas > Kolmogorov-Smirnov)

Este ejemplo muestra cómo utilizar la prueba de *Kolmogorov-Smirnov* para contrastar la hipótesis de normalidad (ajuste a la distribución normal) referida a la variable *salario inicial*:

- En el cuadro de diálogo *Prueba de Kolmogorov-Smirnov* (ver figura 19.5), seleccionar la variable *salario inicial* y trasladarla a la lista **Contrastar variables**.
- Dejar marcada la opción **Normal** del recuadro **Distribución de contraste** para efectuar el ajuste a la distribución normal.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran la tabla 19.5. La tabla ofrece, en primer lugar, el número de casos válidos y los parámetros de la distribución seleccionada, es decir, de la distribución normal (*Media* y *Desviación típica*). A continuación ofrece las diferencias más extremas entre las frecuencias acumuladas empíricas y teóricas (la más grande de las positivas, la más pequeña de las negativas y la más grande de las dos en valor absoluto). Por último, ofrece el estadístico de K-S ($Z = 5{,}484$) y su nivel crítico (*Significación* *asintótica bilateral* = 0,000). Puesto que el valor del nivel crítico es muy pequeño (menor que 0,05), rechazanos la hipótesis de normalidad y concluimos que las puntuaciones de la variable *salario inicial* no se ajustan a una distribución normal.

Tabla 19.6. Prueba de Kolmogorov-Smirnov para una muestra.

		Salario inicial
N		474
Parámetros normales[a,b]	Media	$17,016.09
	Desviación típica	$7,870.64
Diferencias más extremas	Absoluta	,252
	Positiva	,252
	Negativa	-,170
Z de Kolmogorov-Smirnov		5,484
Sig. asintót. (bilateral)		,000

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

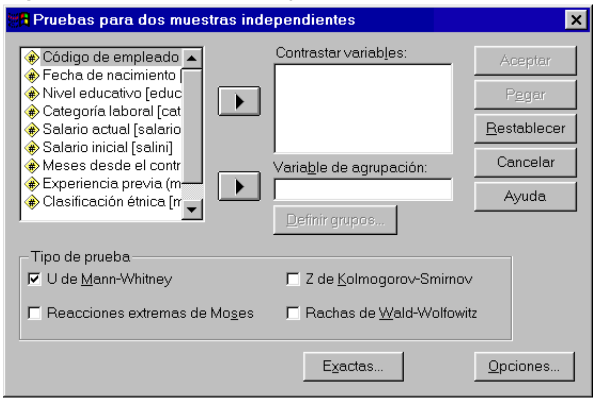
Pruebas para dos muestras independientes

Este procedimiento contiene varias pruebas no paramétricas, todas ellas diseñadas para analizar datos provenientes de diseños con una variable independiente categórica (con dos niveles que definen dos grupos o muestras) y una variable dependiente cuantitativa al menos ordinal (en la cual interesa comparar los dos grupos o muestras).

El procedimiento incluye cuatro pruebas: la prueba *U* de Mann-Whitney, la prueba de Kolmogorov-Smirnov para dos muestras, la prueba de *reacciones extremas* de Moses y la prueba de las *rachas* de Wald-Wolfowitz. Para obtener cualquiera de ellas:

▶ Seleccionar la opción **Pruebas no paramétricas > 2 muestras independientes** del menú **Analizar** para acceder al cuadro de diálogo *Pruebas para dos muestras independientes* que muestra la figura 19.6.

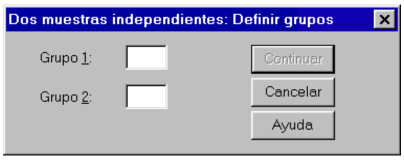
Figura 19.6. Cuadro de diálogo *Pruebas para dos muestras independientes*.



La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para obtener cualquiera de las pruebas no paramétricas incluidas en el procedimiento (puede seleccionarse más de una simultáneamente):

- Seleccionar la variable en la que interesa comparar los grupos y trasladarla a la lista **Contrastar variables**. Si se selecciona más de una variable, el SPSS ofrece un contraste por cada variable seleccionada.
- Seleccionar la variable que define los dos grupos (muestras) que interesa comparar y trasladarla al cuadro **Variable de agrupación**.
- Pinchar el botón **Definir grupos...** para acceder al subcuadro de diálogo *Dos muestras independientes: Definir grupos* que muestra la figura 19.7, el cual permite indicar cuáles son los dos códigos de la variable de agrupación que corresponden a los grupos (muestras) que interesa comparar.

Figura 19.7. Subcuadro de diálogo *Definir grupos*.



- En el recuadro **Tipo de prueba**, marcar la opción u opciones correspondientes a las pruebas que se desea obtener. Conviene tener en cuenta que no todas ellas permiten contrastar la misma hipótesis.

El botón **Opciones...** conduce a un subcuadro de diálogo (ver figura 19.2) que permite obtener algunos estadísticos descriptivos y controlar el tratamiento de los valores perdidos.

Prueba *U* de Mann-Whitney

La prueba *U* de Mann-Whitney[*] es una excelente alternativa a la prueba *t* sobre diferencia de medias cuando: 1) no se cumplen los supuestos en los que se basa la prueba *t* (normalidad y homocedasticidad), o 2) no es apropiado utilizar la prueba *t* porque el nivel de medida de los datos es ordinal.

Consideremos dos muestras independientes: Y_1, de tamaño n_1, e Y_2, de tamaño n_2, extraídas de la misma población o de dos poblaciones idénticas. Si mezclamos las $n_1+n_2 = n$ observa-

[*] El procedimiento que en este apartado estamos llamando *prueba U de Mann-Whitney* fue originalmente propuesto por Wilcoxon (1945) para el caso de tamaños muestrales iguales ($n_1 = n_2$). Festinger (1946) desarrolló independientemente un procedimiento equivalente al de Wilcoxon. Pero fueron Mann y Whitney (1947) los primeros en extender el procedimiento al caso de tamaños muestrales desiguales y los primeros también en proporcionar tablas para poder utilizar el procedimiento con muestras pequeñas. Fueron precisamente las aportaciones de Mann y Whitney las que más contribuyeron a la divulgación del procedimiento; de ahí que, generalmente, sea conocido como prueba de Mann-Whitney. Sin embargo, en algunos contextos este procedimiento todavía puede encontrarse con la denominación de prueba de Wilcoxon-Mann-Whitney, o prueba de Wilcoxon para muestras independientes (la cual no debe ser confundida con la prueba de Wilcoxon para dos muestras relacionadas que se explica más adelante en este mismo capítulo).

ciones y, como si se tratara de una sola muestra, asignamos rangos R_i a las n puntuaciones (un 1 a la más pequeña, un 2 a la más pequeña de las restantes, ..., un n a la más grande; resolviendo los empates asignando el rango promedio), tendremos n_1 rangos R_{i1} (los n_1 rangos correspondientes a las observaciones de la muestra Y_1) y n_2 rangos R_{i2} (los n_2 rangos correspondientes a las observaciones de la muestra Y_2).

Consideremos ahora, entre los múltiples estadísticos que podríamos definir en una situación como ésta, los estadísticos S_1 = "suma de los rangos asignados a la muestra 1" y S_2 = "suma de los rangos asignados a la muestra 2". El estadístico U adopta la siguiente forma en cada grupo:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - S_1 \qquad \text{y} \qquad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - S_2$$

Puesto que suponemos que las dos muestras se han extraído de dos poblaciones idénticas, cabe esperar que U_1 y U_2 sean aproximadamente iguales (excepto en la cantidad atribuible a las fluctuaciones propias del azar muestral). Si U_1 y U_2 son muy distintos, existirá cierta evidencia de que las muestras proceden de poblaciones distintas. Por tanto, la hipótesis nula de que ambos promedios poblacionales son iguales podría rechazarse si U_1 (o U_2) es demasiado grande o demasiado pequeño.

Para determinar esto último, podemos basar nuestra decisión en la probabilidad concreta asociada al estadístico U:

$$U = U_1 \quad \text{si } U_1 < n_1 n_2/2$$
$$U = U_2 \quad \text{si } U_1 > n_1 n_2/2$$

Con muestras pequeñas ($n \leq 30$) el SPSS ofrece el nivel crítico bilateral exacto asociado al estadístico U, el cual se obtiene multiplicando por 2 la probabilidad de obtener valores menores o iguales que U (esta probabilidad se calcula utilizando el algoritmo de Dineen y Blakesley, 1973).

Con muestras grandes ($n > 30$), el SPSS ofrece una tipificación* del estadístico U (incluyendo corrección por empates) que se distribuye aproximadamente $N(0, 1)$:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)}\left(\frac{n^3 - n}{12} - \sum_i^k \frac{t_i^3 - t_i}{12}\right)}}$$

(k se refiere al número de rangos distintos en los que existen empates y t_i al número de puntuaciones empatadas en el rango i). El nivel crítico bilateral se obtiene multiplicando por 2 la probabilidad de obtener valores menores o iguales que Z.

* Existen diferentes versiones de los estadísticos U y Z (ver, por ejemplo, San Martín y Pardo, 1989, pág. 126; o Marascuilo y McSweeney, 1977, págs. 267-278), pero todas ellas conducen al mismo resultado.

Prueba de *reacciones extremas* de Moses

Esta prueba sirve para estudiar si existe diferencia en el grado de dispersión o variabilidad de dos distribuciones.

Aunque hasta ahora hemos hablado de la heterogeneidad de varianzas como de algo relacionado con la prueba T sobre diferencia de medias y, por tanto, como algo poco deseable, lo cierto es que la heterogeneidad de varianzas puede constituir, ella misma, un resultado experimental relevante. Esto significa que, en ocasiones, el estudio de la variabilidad puede ser un fin en sí misma y no sólo un paso previo para la comparación de medias (ver, por ejemplo, Bryk y Raudenbush, 1988).

Supongamos que deseamos evaluar el nivel de desarrollo cognitivo alcanzado por dos grupos de niños que han seguido programas educativos distintos. Si estamos interesados simplemente en constatar cuál de los dos grupos ha alcanzado, en promedio, mayor nivel de desarrollo, podemos limitarnos a comparar los promedios de ambos grupos con alguno de los procedimientos (paramétricos o no paramétricos) ya estudiados. Pero esta forma de proceder pasaría por alto una cuestión de gran importancia: podría ocurrir que uno de los métodos educativos consiguiera incrementar el nivel de desarrollo de los niños de forma generalizada (todos los niños mejoran su nivel de desarrollo) y que el otro método educativo consiguiera el mismo objetivo con sólo unos pocos niños, aunque de forma más marcada, o podría ocurrir que consiguiera incrementar mucho el nivel de desarrollo de unos niños y muy poco el de otros (reacciones extremas). Estas diferencias entre métodos no quedarían reflejadas en las medias, pero sí en la variabilidad, por lo que sólo acompañando el contraste de medias con un contraste de varianzas podríamos obtener información real sobre lo que está ocurriendo.

Existen diferentes procedimientos para contrastar la hipótesis de que dos varianzas poblacionales son iguales. Ya hemos estudiado (ver capítulo 11 sobre *Análisis exploratorio*) uno de los más utilizados, debido a Levene (1960); pero es un procedimiento paramétrico. Moses (1952) ha diseñado un método no paramétrico que puede utilizarse con variables ordinales.

Consideremos dos muestras (c = *control* y e = *experimental*) extraídas aleatoriamente de la misma población o de dos poblaciones idénticas. Para obtener el estadístico de Moses se comienza ordenando las $n = n_c + n_e$ observaciones de forma ascendente y asignándoles, como si se tratara de una única muestra, rangos de 1 a n: un 1 a la más pequeña, un 2 a la más pequeña de las restantes, etc. (los empates se resuelven asignando el rango medio). A continuación se calcula la *amplitud del grupo control* (A_c) restando los rangos correspondientes al valor más grande y más pequeño de ese grupo y sumando 1 a esa diferencia; el resultado se redondea al entero más próximo (el SPSS considera que el grupo control es el grupo con el código menor).

Dado que la amplitud es una medida de dispersión muy inestable, Moses sugiere utilizar al *amplitud recortada* (A_r). Para ello, se fija un valor pequeño (r) y se calcula la amplitud tras descartar r valores del grupo control por arriba y por abajo ($2r$ valores en total). La amplitud recortada se obtiene restando los rangos correspondientes al valor más grande y al más pequeño del grupo control tras eliminar del cómputo los r valores más grandes y los r valores más pequeños de ese grupo; y, por supuesto, sumando 1 a esa diferencia y redondeando al entero más próximo.

Es evidente que A_r no puede ser menor que $n_c–2r$ (ni mayor que $n–2r$). Además, si en el grupo experimental se han producido reacciones extremas, la amplitud del grupo control tenderá a su valor mínimo, pues habrá pocas observaciones del grupo experimental entremezcladas con las del control. Por tanto, podría resultar muy informativo conocer la probabilidad asociada a los valores A_r que superen en alguna cantidad el valor n_c–$2r$. Si llamamos s a la cantidad en que un determinado valor observado de A_r supera a n_c–$2r$, podemos obtener la probabilidad de encontrar valores A_r como el observado o menores mediante:

$$P(A_s \leq n_c - 2r + s) = \frac{\sum_{i=0}^{s} \left[\binom{i + n_c - 2r - 2}{i} \binom{n_e + 2r + 1 + - i}{n_e - i} \right]}{\binom{n}{n_c}}$$

El SPSS calcula esta probabilidad tanto para $r = 0$ como para $r = 0{,}005n$ (en este último caso, si $r < 0$, se toma 1). Si esa probabilidad es demasiado pequeña, podremos rechazar la hipótesis de que ambas muestras proceden de poblaciones con la misma amplitud.

Prueba de Kolmogorov-Smirnov para dos muestras

Esta prueba sirve para contrastar la hipótesis de que dos muestras proceden de la misma población. Para ello, compara las funciones de distribución (funciones de probabilidad acumuladas) de ambas muestras: $F_1(X_i)$ y $F_2(X_i)$. A diferencia de lo que ocurre con la prueba U de Mann-Whitney, que permite comparar dos promedios poblacionales, la prueba de Kolmogorov-Smirnov es sensible a cualquier tipo de diferencia entre las dos distribuciones (tendencia central, simetría, variabilidad, etc.).

Para obtener las funciones de distribución de las dos muestras se comienza asignando rangos a los valores de X_i. Esta asignación de rangos se realiza de forma separada para cada muestra y los empates se resuelven asignando el rango promedio a las puntuaciones empatadas.

Tras asignar rangos a los valores de ambas muestras, la función de distribución empírica para cada valor de X_i se obtiene, en cada muestra, de la siguiente manera: $F_j(X_i) = i \,/\, n_j$ (donde i se refiere al rango correspondiente a cada observación). A continuación se obtienen las diferencias $D_i = F_1(X_i) - F_2(X_i)$, donde $F_1(X_i)$ se refiere a la función de distribución de la muestra de mayor tamaño. Una vez obtenidas las diferencias D_i, la hipótesis de las dos muestras proceden de la misma población se pone a prueba utilizando una tipificación de la diferencia más grande en valor absoluto (Smirnov, 1939, 1948):

$$Z_{K-S} = máx_i \, |D_i| \, \sqrt{(n_1 n_2)/(n_1 + n_2)}$$

Este estadístico Z se distribuye según el modelo de probabilidad normal $N(0, 1)$. El SPSS utiliza el método de Smirnov (1948) para obtener las probabilidades concretas asociadas a los valores del estadístico Z. Este método difiere del estándar (basado en las probabilidades de la curva normal estandarizada), pero es equivalente. Si la probabilidad de obtener una diferencia tan grade como la observada es muy pequeña (generalmente, menor que 0,05), podremos rechazar la hipótesis de que ambas muestras proceden de la misma población.

Prueba de las *rachas* de Wald-Wolfowitz

La prueba de las *rachas* para dos muestras independientes (Wald y Wolfowitz, 1940) es similar a la prueba de las *rachas* para una muestra ya estudiada en este mismo capítulo. Referida a dos muestras independientes, permite contrastar la hipótesis de que ambas muestras proceden de la misma población. Al igual que la prueba de Kolmogorov-Smirnov, es sensible no sólo a diferencias entre los promedios poblaciones, sino a diferencias en variabilidad, simetría, etc.

Para obtener el número de rachas, se ordenan de menor a mayor las $n = n_1 + n_2$ observaciones de ambas muestras como si se tratara de una sola muestra. Ordenadas las puntuaciones, el número de rachas (R) se obtiene contando el número de secuencias de observaciones pertenecientes al mismo grupo. Si existen empates entre observaciones de muestras distintas, el SPSS calcula tanto el número mínimo de rachas como el máximo.

Si las dos muestras proceden de la misma población, las observaciones de ambas muestras estarán entremezcladas y el número de rachas será alto. Por el contrario, si las muestras proceden de poblaciones distintas, una de ellas tendrá valores más altos que la otra y, al ordenar las observaciones, no estarán tan entremezcladas. Así pues, un número alto de rachas indica que las muestras proceden de la misma población y un número bajo de rachas indica que las muestras proceden de poblaciones distintas.

Para decidir cuándo el número de rachas encontrado es lo bastante pequeño como para rechazar la hipótesis de que las muestras proceden de la misma población, el SPSS utiliza dos estrategias distintas dependiendo del tamaño de las muestras. Si $n > 30$, utiliza la aproximación normal (ver, en este mismo capítulo, el estadístico Z descrito en el apartado *Prueba de las rachas*); pero a diferencia de lo que ocurre con el estadístico Z para una muestra, aquí se utiliza un nivel crítico unilateral: la probabilidad de obtener un número de rachas igual o menor que el obtenido.

Si $n \leq 30$, el SPSS ofrece el nivel crítico unilateral exacto. Para ello, si el número de rachas R es par, utiliza la siguiente ecuación:

$$P(r \leq R) = \frac{2}{\binom{n}{n_1}} \sum_{r=2}^{R} \binom{n_1 - 1}{r/2 - 1} \binom{n_2 - 1}{r/2 - 1}$$

Y si el número de rachas R es impar ($k = 2r - 1$):

$$P(r \leq R) = \frac{1}{\binom{n}{n_1}} \sum_{r=2}^{R} \binom{n_1 - 1}{k - 1} \binom{n_2 - 1}{k - 2} + \binom{n_1 - 1}{k - 2} \binom{n_2 - 1}{k - 1}$$

En ambos casos se está haciendo referencia a la probabilidad de obtener un número de rachas igual o menor que el encontrado. Se rechaza la hipótesis nula de que las muestras proceden de la misma población cuando la probabilidad obtenida es menor que, generalmente, 0,05.

Ejemplo (Pruebas no paramétricas > Dos muestras independientes)

Este ejemplo muestra cómo obtener e interpretar todas las pruebas incluidas en el procedimiento **Pruebas no paramétricas** > **Dos muestras independientes**:

- En el cuadro de diálogo *Pruebas para dos muestras independientes* (ver tabla 19.6), seleccionar la variable *salini* (salario inicial) y trasladarla a la lista **Contrastar variables**.
- Seleccionar la variable *minoría* (clasificación étnica) y trasladarla al cuadro **Variable de agrupación**.
- Pinchar el botón **Definir grupos...** para acceder al subcuadro de diálogo *Dos muestras independientes: Definir grupos* que muestra la figura 19.7, e introducir los códigos 0 y 1 (que son los códigos que definen los dos grupos de la variable *minoría*).
- En el recuadro **Tipo de prueba**, marcar las cuatro opciones disponibles.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran las tablas 19.7 a la 19.11. La primera de ellas (tabla 19.7) ofrece el tamaño de cada grupo, el rango promedio que resulta de la asignación de rangos a cada grupo y la *suma* de esos rangos.

Tabla 19.7. Rangos.

Salario inicial

Clasificación étnica	N	Rango promedio	Suma de rangos
No	370	249,14	92180,50
Sí	104	196,10	20394,50
Total	474		

La tabla 19.8 ofrece el estadístico *U* de *Mann-Whitney* (también ofrece el estadístico *W* de *Wilcoxon*, que es una versión equivalente del estadístico *U*; ver Pardo y San Martín, 1994, págs. 424-427). La tipificación de ambos vale $Z = -3{,}495$. Y el nivel crítico bilateral (*Significación asintótica bilateral*) vale 0,000. Por tanto, podemos rechazar la hipótesis de igualdad de promedios y concluir que los grupos definidos por la variable *minoría* proceden de poblaciones con distinto promedio.

Tabla 19.8. Prueba de *Mann-Whitney*.

	Salario inicial
U de Mann-Whitney	14934,500
W de Wilcoxon	20394,500
Z	-3,495
Sig. asintót. (bilateral)	,000

La tabla 19.9. contiene la prueba de reacciones extremas de Moses. Por supuesto, la variable *minoría* no parece del todo apropiada para utilizar la prueba de Moses, pero, puesto que se trata de un ejemplo, nos sirve para poder interpretar la información que ofrece el *Visor*. La tabla recoge, en primer lugar, la amplitud del grupo control ($N = 467$) y la probabilidad de obtener una amplitud como esa o menor (*Significación unilateral* = 0,000). A continuación muestra la amplitud recortada ($N = 434$) y la probabilidad de obtener una amplitud como esa o menor (*Signi-*

ficación unilateral = 0,990). Puesto que el nivel crítico es mayor que 0,05, podemos considerar que no se han producido reacciones extremas. La última línea recoge el valor de *r*, es decir, el número de casos eliminados por arriba y por abajo para obtener la amplitud recortada (este número es, recordemos, el 5 por ciento del tamaño muestral).

Tabla 19.9. Prueba de *Moses*.

		Salario inicial
Amplitud observada del grupo control	N	467
	Sig. (unilateral)	,000
Amplitud recortada del grupo control	N	434
	Sig. (unilateral)	,990
Valores atípicos recortados de cada extremo		18

La tabla 19.10 ofrece la prueba de *Kolmogorov-Smirnov*. En primer lugar aparecen las diferencias más extremas (*absoluta*, *positiva* y *negativa*) entre las funciones de distribución de ambas muestras. Y a continuación se muestra el resultado de la tipificación de la diferencia más extrema en valor absoluto (*Z de Kolmogorov-Smirnov* = 2,134) junto con el nivel crítico bilateral (*Significación asintótica bilateral* = 0). Puesto que el nivel crítico es menor que 0,05, podemos rechazar la hipótesis de igualdad de distribuciones y concluir que los dos grupos comparados difieren significativamente en *salario inicial.*

Tabla 19.10. Prueba de *Kolmogorov-Smirnov*.

		Salario inicial
Diferencias más extremas	Absoluta	,237
	Positiva	,000
	Negativa	-,237
Z de Kolmogorov-Smirnov		2,134
Sig. asintót. (bilateral)		,000

La tabla 19.11 recoge el resultado de la prueba de las *rachas*. La tabla ofrece el número mínimo y máximo de rachas (dependiendo del tratamiento que se dé a los empates), el valor del estadístico *Z* en cada supuesto y el nivel crítico unilateral (*Significación asintótica unilateral*). La presencia de 25 empates hace que el resultado obtenido en ambas situaciones sea muy diferente. En estos casos, puede recurrirse a otras pruebas para tomar una decisión.

Tabla 19.11. Prueba de las *rachas* de *Wald-Wolfowitz*.

Salario inicial

	Número de rachas	Z	Sig. asintót. (unilateral)
Mínimo posible	40[a]	-16,576	,000
Máximo posible	200[a]	4,923	1,000

a. Hay 25 empates inter-grupos que implican 348 casos.

Pruebas para varias muestras independientes

Este procedimiento contiene varias pruebas no paramétricas, todas ellas diseñadas para analizar datos provenientes de diseños con una variable independiente categórica (con más de dos niveles que definen más de dos grupos o muestras) y una variable dependiente cuantitativa al menos ordinal en la cual interesa comparar las muestras. El procedimiento incluye tres pruebas: la prueba *H* de Kruskal-Wallis, la prueba de la *mediana* y la prueba de Jonckheere-Terpstra (ésta última sólo se incluye en el módulo *Pruebas exactas*). Para obtener cualquiera de ellas:

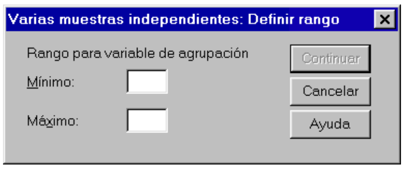
- ▶ Seleccionar la opción **Pruebas no paramétricas** > **K muestras independientes** del menú **Analizar** para acceder al cuadro de diálogo *Pruebas para varias muestras independientes* que muestra la figura 19.6.

Figura 19.8. Cuadro de diálogo *Pruebas para varias muestras independientes*.

La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para obtener cualquiera de las pruebas no paramétricas incluidas en el procedimiento (puede seleccionarse más de una simultáneamente):

- ▶ Seleccionar la variable en la que interesa comparar los grupos y trasladarla a la lista **Contrastar variables**. Si se selecciona más de una variable, el SPSS ofrece un contraste por cada variable seleccionada.
- ▶ Seleccionar la variable que define los grupos (muestras) que interesa comparar y trasladarla al cuadro **Variable de agrupación**.
- ▶ Pinchar el botón **Definir grupos...** para acceder al subcuadro de diálogo *Varias muestras independientes: Definir grupos* que muestra la figura 19.9.

Figura 19.9. Subcuadro de diálogo *Varias muestras independientes: Definir grupos*.



Este subcuadro de diálogo permite indicar cuáles son los dos códigos de la variable de agrupación que corresponden a los grupos (muestras) que interesa comparar: se comparan los grupos con los códigos **Mínimo** y **Máximo** y todos los comprendidos entre ellos.

▶ En el recuadro **Tipo de prueba**, marcar la opción u opciones correspondientes a las pruebas que se desea obtener. Conviene tener en cuenta que no todas ellas permiten contrastar la misma hipótesis.

El botón **Opciones...** conduce a un subcuadro de diálogo idéntico al de la figura 19.2 que permite obtener algunos estadísticos descriptivos y decidir qué tratamiento se desea dar a los valores perdidos.

Prueba *H* de *Kruskal-Wallis*

La prueba de Mann-Whitney para dos muestras independientes fue extendida al caso de más de dos muestras por Kruskal y Wallis (1952). La situación experimental que permite resolver esta prueba es similar a la estudiada a propósito del ANOVA de un factor completamente aleatorizado: J muestras son aleatoria e independientemente extraídas de J poblaciones para averiguar si las J poblaciones son idénticas o alguna de ellas presenta promedios mayores que otra.

Las ventajas fundamentales de esta prueba frente al estadístico F del ANOVA de un factor completamente aleatorizado son dos: 1) no necesita establecer supuestos sobre las poblaciones originales tan exigentes como los del estadístico F (normalidad, homocedasticidad); y 2) permite trabajar con datos ordinales. Por contra, si se cumplen los supuestos en los que se basa el estadístico F, la potencia de éste es mayor que la que es posible alcanzar con el estadístico H de Kruskal-Wallis.

Ahora bien, teniendo en cuenta que en muchas situaciones reales resulta demasiado arriesgado suponer normalidad y homocedasticidad (especialmente si las muestras son pequeñas y/o los tamaños muestrales desiguales), y considerando además que en otras situaciones el nivel de medida de los datos puede no ir más allá del ordinal, la prueba de Kruskal-Wallis* representa una excelente alternativa al ANOVA de un factor completamente aleatorizado.

Consideremos J muestras aleatorias e independientes de tamaños $n_1, n_2, ..., n_J$ extraídas de la misma población o de J poblaciones idénticas. Llamemos n al conjunto total de observaciones: $n = n_1 + n_2 + \cdots + n_J$. Asignemos rangos desde 1 hasta n a ese conjunto de n observaciones como si se tratara de una sola muestra (si existen empates se asigna el promedio de los rangos empatados).

Llamemos R_{ij} a los rangos asignados a las observaciones i de la muestra j. Llamemos R_j a la suma de los rangos asignados a las n_j observaciones de la muestra j. Tendremos:

$$R_j = \sum_i^{n_j} R_{ij} \qquad \text{y} \qquad \bar{R}_j = \frac{R_j}{n_j}$$

* No es infrecuente encontrarse manuales de estadística en los que la prueba H de Kruskal-Wallis aparece con la denominación *análisis de varianza por rangos*.

Obviamente, si la hipótesis nula de que las J poblaciones son idénticas es verdadera, los R_j de las distintas muestras serán parecidos. Siguiendo una lógica similar a la del estadístico U de Mann-Whitney, es posible obtener, tomando como punto de partida la suma de los rangos de cada muestra, un estadístico con distribución muestral conocida y capaz de proporcionarnos información sobre el parecido existente entre las J poblaciones (ver, por ejemplo, San Martín y Pardo, 1989, págs. 225-227)*:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{J} \frac{R_j^2}{n_j} - 3(n+1)$$

Bajo la hipótesis nula de que los J promedios poblacionales son iguales, el estadístico H se distribuye según el modelo de probabilidad *chi-cuadrado*, con J–1 grados de libertad.

Prueba de la *mediana*

La prueba de la mediana es similar a la prueba *chi-cuadrado* ya estudiada en el capítulo 12 sobre *Tablas de contingencia*. La única diferencia entre ambas es que ahora, en lugar de utilizar dos variables categóricas, una de ellas es cuantitativa y se dicotomiza utilizando la mediana (de ahí el nombre de la prueba).

Tenemos, por tanto, una variable categórica que define J muestras de tamaño n_j (Siendo $n = \sum n_j$) y una variable al menos ordinal. El objetivo de la prueba de la mediana es contrastar la hipótesis de que las J muestras proceden de poblaciones con la misma mediana. Para ello, se comienza ordenando todas las observaciones y calculando la mediana total (la mediana de las n observaciones):

$$Mdn = (X_{[n/2]} + X_{[n/2+1]})/2 \qquad \text{si } n \text{ es par}$$
$$Mdn = X_{[(n+1)/2]} \qquad \text{si } n \text{ es impar}$$

(donde $X_{[n]}$ se refiere al valor más grande y $X_{[1]}$ al más pequeño). A continuación se registra, dentro de cada muestra, el número de casos con puntuación igual o menor que la mediana (grupo 1) y el número de casos con valor mayor que la mediana (grupo = 2). Tras esto, se construye una tabla de contingencia bidimensional de tamaño 2×J, con las 2 filas correspondientes a los dos grupos obtenidos al dicotomizar por la mediana y las J columnas correspondientes a las J muestras independientes. Por último, se aplica el estadístico *chi-cuadrado* ya estudiado en el capítulo 12, apartado *Estadísticos > Chi-cuadrado*. Las frecuencias esperadas se obtienen suponiendo que los 2 grupos y las J muestras son independientes.

* En el caso de que existan empates, el SPSS utiliza una corrección que hace el contraste algo más conservador:

$$H' = \frac{H}{1 - \sum_{i=1}^{k} (t_i^3 - t_i) \Big/ (n^3 - n)}$$

donde k se refiere al número de rangos distintos en los que existen empates y t_i al número de valores empatados en cada rango.

Ejemplo (Pruebas no paramétricas > Varias muestras independientes)

Este ejemplo muestra cómo obtener e interpretar las pruebas incluidas en el procedimiento **Pruebas no paramétricas** > **K muestras independientes**:

- ▶ En el cuadro de diálogo *Pruebas para varias muestras independientes* (ver figura 19.8), seleccionar la variable *salini* (salario inicial) y trasladarla a la lista **Contrastar variables**.
- ▶ Seleccionar la variable *catlab* (categoría laboral) y trasladarla al cuadro **Variable de agrupación**.
- ▶ Pulsar el botón **Definir grupos...** para acceder al subcuadro de diálogo *Varias muestras independientes: Definir grupos* que muestra la figura 19.9, e introducir los códigos 1 y 3 para indicar que deseamos comparar los grupos 1 y 3 y todos los comprendidos entre ellos (en este caso, los grupos 1, 2 y 3).
- ▶ En el recuadro **Tipo de prueba**, marcar las opciones **H de Kruskal-Wallis** y **Mediana**.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran las tablas 19.12 a la 19.15.

La primera de ellas (tabla 19.12) ofrece el tamaño de cada grupo y los rangos promedio resultantes de la asignación de rangos a las puntuaciones de los tres grupos.

Tabla 19.12. Rangos.

Salario inicial

Categoría Laboral	N	Rango promedio
Administrativo	363	192,29
Seguridad	27	252,59
Directivo	84	428,04
Total	474	

La tabla 19.13 contiene el estadístico de Kruskal-Wallis (*Chi-cuadrado*), sus grados de libertad (*gl*) y su nivel crítico (*Sig. asintót.*). Puesto que el nivel crítico (0,000) es menor que 0,05, podemos rechazar la hipótesis de igualdad de promedios poblacionales y concluir que las poblaciones comparadas difieren en *salario inicial*.

Tabla. 19.13. Prueba de Kruskal-Wallis.

	Salario inicial
Chi-cuadrado	203,112
gl	2
Sig. asintót.	,000

Aunque existen procedimientos para efectuar comparaciones múltiples tras obtener un estadístico *H* significativo (ver, por ejemplo, Pardo y San Martín, 1998, págs. 437-441), para analizar con el SPSS qué grupos de *categoría laboral* difieren entre sí podemos utilizar la prueba de Mann-Whitney para dos muestras independientes, pero acompañada de la corrección de Bonferroni para controlar la tasa de error (la probabilidad de cometer errores de tipo I). Puesto que con tres grupos de *categoría laboral* necesitamos hacer tres comparaciones dos a dos (1-2, 1-3 y 2-3), la aplicación de la corrección de Bonferroni nos llevará a basar nuestras decisiones en un nivel de significación de 0,05/3 = 0,017. Es decir, consideraremos que dos grupos difieren significativamente cuando el nivel crítico obtenido sea menor de 0,017.

Las tablas 19.14 y 19.15 contienen la información relacionada con la prueba de la *mediana*. La tabla 19.14 muestra el resultado de la dicotomización, es decir, el resultado de clasificar a los sujetos de cada *categoría laboral* por debajo y por encima de la mediana. Como dato curioso, podemos observar que no existen *directivos* que estén por debajo de la mediana del *salario inicial*.

Tabla 19.14. Frecuencias.

Salario inicial

	Categoría Laboral		
	Administrativo	Seguridad	Directivo
> Mediana	112	14	84
<= Mediana	251	13	0

La tabla 19.15 ofrece el tamaño de la muestra, el valor de la mediana, el estadístico *Chi-cuadrado*, sus grados de libertad y su nivel crítico asintótico. Puesto que el nivel crítico es menor que 0,05, podemos rechazar la hipótesis de igualdad de promedios poblacionales y concluir que las poblaciones comparadas difieren en *salario inicial*.

Tabla 19.15. Prueba de la Mediana.

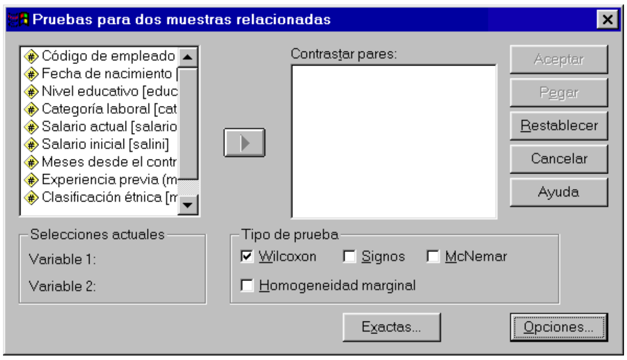
	Salario inicial
N	474
Mediana	$15,000.00
Chi-cuadrado	132,835
gl	2
Sig. asintót.	,000

Pruebas para dos muestras relacionadas

Las pruebas de este apartado permiten analizar datos provenientes de diseños con medidas repetidas. Algunas de estas pruebas (*Wilcoxon* y *Signos*) sirven para contrastar hipótesis sobre igualdad de medianas; la prueba de *McNemar* sirve para contrastar hipótesis sobre igualdad de proporciones (ver el capítulo 12 sobre *Tablas de contingencia* para una descripción de esta prueba). Todas ellas se ajustan a diseños del tipo *antes-después*, pero difieren en el tipo de variables que permiten analizar. Para obtener cualquiera de estas pruebas:

▶ Seleccionar la opción **Pruebas no paramétricas** > **2 muestras relacionadas** del menú **Analizar** para acceder al cuadro de diálogo *Pruebas para dos muestras relacionadas* que recoge la figura 19.10.

Figura 19.10. Cuadro de diálogo *Pruebas para dos muestras relacionadas*.



La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para obtener cualquiera de las pruebas no paramétricas incluidas en el procedimiento (puede seleccionarse más de una simultáneamente):

▶ Seleccionar las variables cuyas medianas o proporciones interesa comparar y trasladarlas a la lista **Contrastar pares**. Puede seleccionarse más de un par de variables: el SPSS ofrece un contraste por cada par seleccionado.

Selecciones actuales. Muestra las variables seleccionadas antes de pulsar el botón flecha.

El botón **Opciones...** conduce a un subcuadro de diálogo idéntico al de la figura 19.2 que permite obtener algunos estadísticos descriptivos y decidir qué tratamiento se desea dar a los valores perdidos.

Prueba de Wilcoxon

Supongamos que tomamos dos medidas (X_i e Y_i) a un grupo de m sujetos y que calculamos las diferencias en valor absoluto entre las dos puntuaciones de cada par:

$$D_i = |X_i - Y_i| \qquad (i = 1, 2, ..., m)$$

Desechemos las D_i nulas y consideremos únicamente las n diferencias D_i no nulas ($n \leq m$). Asignemos *rangos* (R_i) desde 1 hasta n a esas D_i no nulas: el rango 1 a la D_i más pequeña, el rango 2 a la D_i más pequeña de las restantes, ..., el rango n a la D_i más grande (si existen empates, se resuelven asignando el promedio de los rangos). Sumemos ahora, por un lado, los R_i^+, es decir, los rangos correspondientes a las D_i con $X_i > Y_i$, y llamemos S_+ a esta suma; y sumemos, por otro lado, los R_i^-, es decir, los rangos correspondientes a las D_i con $X_i < Y_i$, y llamemos S_- a esta otra suma. Si suponemos que las puntuaciones X_i e Y_i proceden de poblaciones con la misma mediana ($Mdn_x = Mdn_y$), debemos esperar que:

$$P(X_i < Y_i) = P(X_i > Y_i)$$

por lo que, si la hipótesis H_0: $Mdn_x = Mdn_y$ es verdadera, en una muestra aleatoria de n observaciones debemos encontrar aproximadamente tantos valores $X_i > Y_i$ como valores $X_i < Y_i$ (salvando, por supuesto, las fluctuaciones atribuibles al azar muestral). Pero, además, si la distribución de las diferencias es *simétrica* (lo cual exige escala de intervalo o razón), las D_i positivas se alejarán de cero *en igual medida* que las D_i negativas, de donde es fácil deducir que:

$$S_+ = \sum R_i^+ \approx S_- = \sum R_i^-$$

Es decir, si $Mdn_x = Mdn$ y la distribución de las diferencias D_i es simétrica, S_+ y S_- tomarán valores parecidos. Por tanto, una fuerte discrepancia entre S_+ y S_- nos hará dudar de la veracidad de H_0. De modo que podemos utilizar los valores S_+ y S_- para obtener información sobre la hipótesis H_0: $Mdn_x = Mdny$ (Wilcoxon, 1945, 1949).

Con tamaños muestrales pequeños no resulta complicado obtener la distribución exacta del estadístico S_+ (o S_-). Pero es más rápido obtener una tipificación de S (S se refiere al menor de S_+ y S_-) cuya distribución se aproxima, conforme el tamaño muestral va aumentando, al modelo de probabilidad normal $N(0, 1)$:

$$Z = \frac{S - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24} - \displaystyle\sum_{i=1}^{k} \dfrac{t_i^3 - t_i}{48}}}$$

(k se refiere al número rangos distintos en los que existen empates y t_i al número de puntuaciones empatadas en el rango i). El SPSS ofrece el nivel crítico bilateral resultante de multiplicar por 2 la probabilidad de obtener valores menores o iguales que Z.

Prueba de los *signos*

La prueba de los signos guarda una muy estrecha relación con la prueba *binomial* ya estudiada en este mismo capítulo. Al igual que la prueba de Wilcoxon, la prueba de los signos permite contrastar la hipótesis de igualdad entre dos medianas poblacionales. Pero, mientras la prueba de Wilcoxon aprovecha la información ordinal de los datos (aunque exige nivel de medida de intervalo o razón), la prueba de los signos sólo aprovecha de los datos sus propiedades nominales (aunque exige nivel de medida al menos ordinal).

Supongamos que tomamos dos medidas (X_i e Y_i) a un grupo de m sujetos y que calculamos las diferencias:

$$D_i = |X_i - Y_i| \qquad (i = 1, 2, ..., m)$$

entre las dos puntuaciones de cada par. Desechemos las D_i nulas y consideremos únicamente las n diferencias D_i no nulas ($n \leq m$). Si suponemos que las puntuaciones X_i e Y_i proceden de poblaciones con la misma mediana ($Mdn_x = Mdn_y$), debe verificarse que:

$$P(X_i < Y_i) = P(X_i > Y_i) = 0{,}5$$

de modo que, si la hipótesis H_0: $Mdn_x = Mdn_y$ es verdadera, en una muestra aleatoria de n observaciones debemos encontrar aproximadamente tantos valores $X_i > Y_i$ como valores $X_i < Y_i$, es decir, aproximadamente tantas diferencias D_i positivas como negativas (salvando, por supuesto, las fluctuaciones atribuibles al azar propio del proceso de muestreo). Bajo estas condiciones, las variables:

$$n_+ = \textit{número de signos positivos}$$

$$n_- = \textit{número de signos negativos}$$

se distribuyen según el modelo binomial con parámetros n y $\pi = 0{,}50$. De modo que podemos utilizar la distribución binomial para conocer las probabilidades asociadas a n_+ y n_- y, basándonos en ellas, contrastar la hipótesis H_0: $Mdn_x = Mdn_y$.

Si $n \leq 25$, el SPSS toma el valor $k = \min(n_+, n_-)$ y, utilizando las probabilidades de la distribución binomial, calcula el nivel crítico bilateral resultante de multiplicar por 2 la probabilidad de obtener valores iguales o menores que k.

Si $n > 25$, el SPSS tipifica el valor de k (utilizando *corrección por continuidad*) y ofrece el nivel crítico resultante de multiplicar por 2 la probabilidad de encontrar valores iguales o menores que Z:

$$Z = \frac{k + 0{,}5 - n/2}{0{,}5\sqrt{n}}$$

Ejemplo (Pruebas no paramétricas > Dos muestras relacionadas)

Este ejemplo muestra cómo obtener e interpretar las pruebas incluidas en el procedimiento **Pruebas no paramétricas** > **2 muestras relacionadas**:

- ▶ En el cuadro de diálogo *Pruebas para dos muestras relacionadas* (ver figura 19.10), seleccionar las variables *salini* (salario inicial) y *salario* (salario actual) y trasladarlas a la lista **Contrastar pares**.
- ▶ En el recuadro **Tipo de prueba**, marcar las opciones **Signos** y **Wilcoxon** (para la prueba de **McNemar**, consultar el capítulo 12 sobre *Tablas de contingencia*).

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran las tablas 19.16 a la 19.20. La tabla 19.16 ofrece algunos estadísticos descriptivos para las dos variables seleccionadas: el número de casos válidos en ambas variables, la media, la desviación típica, el valor más pequeño, el más grande y los cuartiles.

Tabla 19.16. Estadísticos.

	N	Media	Desviación típica	Mínimo	Máximo	Percentiles 25	Percentiles 50 (Mediana)	Percentiles 75
Salario actual	474	$34,419.57	$17,075.66	$15,750	$135,000	$24,000.00	$28,875.00	$37,162.50
Salario inicial	474	$17,016.09	$7,870.64	$9,000	$79,980	$12,450.00	$15,000.00	$17,617.50

Las dos tablas siguientes contienen información relacionada con la prueba de *Wilcoxon*. La tabla 19.17 ofrece el número, media y suma de los rangos negativos y de los rangos positivos. Las notas a pie de tabla permiten conocer el significado de los rangos positivos y negativos. También ofrece el número de empates (casos que no son incluidos en el análisis) y el número total de sujetos.

Tabla 19.17. Rangos.

Salario inicial - Salario actual

	N	Rango promedio	Suma de rangos
Rangos negativos	474[a]	237,50	112575,00
Rangos positivos	0[b]	,00	,00
Empates	0[c]		
Total	474		

a. Salario inicial < Salario actual

b. Salario inicial > Salario actual

c. Salario actual = Salario inicial

La tabla 19.18 muestra el estadístico de *Wilcoxon* (*Z*) y su nivel crítico bilateral (*Sig. asintót. bilateral*). Puesto que el valor del nivel crítico (0,000) es menor que 0,05, podemos rechazar la hipótesis de igualdad de promedios y concluir que las variables comparadas (*salario inicial* y *salario actual*) difieren significativamente.

Tabla 19.18. Prueba de Wilcoxon.

	Salario inicial - Salario actual
Z	-18,865[a]
Sig. asintót. (bilateral)	,000

a. Basado en los rangos positivos.

Las tablas 19.19 y 19.20 contienen la información relacionada con la prueba de los *signos*. La tabla 19.19 muestra las diferencias negativas, las positivas y los empates. Las notas a pie de tabla permiten saber qué diferencias se están considerando negativas y cuáles positivas.

Tabla 19.19. Frecuencias.

Salario inicial - Salario actual

	N
Diferencias negativas [a]	474
Diferencias positivas [b]	0
Empates [c]	0
Total	474

a. Salario inicial < Salario actual

b. Salario inicial > Salario actual

c. Salario actual = Salario inicial

La tabla 19.20 ofrece el estadístico *Z* (pues el tamaño muestral es mayor de 25) y su correspondiente nivel crítico bilateral (*Sig. asintót. bilateral).* Puesto que el valor del nivel crítico

(0,000) es menor que 0,05, podemos rechazar la hipótesis de igualdad de promedios y concluir que las variables comparadas (*salario inicial* y *salario actual*) difieren significativamente.

Tabla 19.20. Prueba de los signos.

	Salario inicial - Salario actual
Z	-21,726
Sig. asintót. (bilateral)	,000

Pruebas para varias muestras relacionadas

Las pruebas agrupadas en este apartado permiten analizar datos provenientes de diseños con medidas repetidas. La prueba de Friedman y el coeficiente de concordancia W de Kendall sirven para estudiar J medidas ordinales; la prueba de Cochran permite contrastar la hipótesis de igualdad de proporciones con J variables dicotómicas. Para obtener cualquiera de ellas:

- Seleccionar la opción **Pruebas no paramétricas** > **K muestras relacionadas** del menú **Analizar** para acceder al cuadro de diálogo *Pruebas para varias muestras relacionadas* que recoge la figura 19.11.

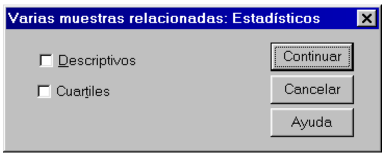
Figura 19.11. Cuadro de diálogo *Pruebas para varias muestras relacionadas*.

La lista de variables del archivo de datos ofrece un listado de todas las variables con formato numérico. Para obtener cualquiera de las pruebas no paramétricas incluidas en el procedimiento (puede seleccionarse más de una simultáneamente, pero debe tenerse presente que no tiene sentido mezclar la prueba de Cochran con las de Friedman y Kendal):

- Seleccionar las variables cuyas medianas o proporciones interesa comparar y trasladarlas a la lista **Contrastar variables**.

El botón **Estadísticos...** da acceso al subcuadro de diálogo *Varias muestras relacionadas: Estadísticos* que muestra la figura 19.12. Este subcuadro de diálogo permite obtener varios estadísticos descriptivos (tamaño muestral, media, desviación típica, valor mínimo y máximo) y los cuartiles.

Figura 19.12. Subcuadro de diálogo *Varias muestras relacionadas: Estadísticos*.



Prueba de Friedman

La prueba de Friedman (1937) sirve para comparar J promedios poblacionales cuando se trabaja con muestras relacionadas. La situación experimental que permite resolver esta prueba es similar a la estudiada a propósito del ANOVA de un factor con *medidas repetidas*: a n sujetos (o a n bloques, cada uno de tamaño J) se le aplican J tratamientos o se le toman J medidas con intención de averiguar si los promedios de esos J tratamientos o medidas son o no iguales.

Las ventajas de esta prueba frente al estadístico F del ANOVA son las mismas que hemos señalado a propósito del estadístico H de Kruskal-Wallis: no es necesario establecer los supuestos tan exigentes del ANOVA (normalidad, igualdad de varianzas) y permite trabajar con datos ordinales. La prueba de Friedman, por tanto, constituye una alternativa al estadístico F cuando no se cumplen los supuestos paramétricos del ANOVA o el nivel de medida de los datos es ordinal.

El diseño está formado por J muestras o tratamientos relacionados y por una muestra aleatoria de n sujetos o bloques independientes entre sí e independientes de los tratamientos. Las puntuaciones originales deben ser transformadas en rangos R_{ij}. Esos rangos se asignan independientemente para cada sujeto o bloque; es decir, se asignan rangos de 1 a J a las observaciones del sujeto o bloque 1; lo mismo con el bloque 2; y lo mismo con el resto de los bloques por separado.

Los rangos asignados a cada sujeto o bloque suman, en todos los casos, $J(J+1)/2$ (pues en cada sujeto o bloque estamos asignando rangos desde 1 a J). Llamaremos R_{ij} al rango asignado al sujeto o bloque i en el tratamiento o muestra j. Y R_j a la suma de los rangos asignados a las n observaciones de la muestra j:

$$R_j = \sum_i^n R_{ij} \qquad \Rightarrow \qquad \bar{R}_j = \frac{R_j}{n}$$

Obviamente, si los promedios poblacionales son iguales, los R_j serán parecidos. Tomando como punto de partida estas sumas de rangos, Friedman (1937) ha diseñado un estadístico con distribución muestral conocida capaz de proporcionarnos información sobre el parecido existente entre las J poblaciones (ver, por ejemplo, Pardo y San Martín, 1998, págs. 441-445):

$$X_r^2 = \frac{12}{nJ(J+1)} \sum_j R_j^2 - 3n(J+1)$$

El estadístico de Friedman se distribuye según el modelo de probabilidad *chi-cuadrado* con J–1 grados de libertad.

Coeficiente de concordancia W de Kendall.

El coeficiente de concordancia W (obtenido independientemente por Kendall y Babington-Smith, 1939, y por Wallis, 1939) sirve para estudiar la relación (acuerdo, concordancia) entre $J > 2$ conjuntos de rangos. La necesidad de estudiar la relación entre J conjuntos de rangos se presenta con cierta frecuencia en diferentes áreas de conocimiento. Tales situaciones se producen, por ejemplo, cuando una muestra aleatoria de n sujetos u objetos es clasificada según J características; o cuando J jueces evalúan, ordenan o clasifican una muestra de n sujetos u objetos según una característica. Cualquiera que sea la forma de obtener ese conjunto de J rangos, podemos llamar R_{ij} al rango que corresponde al sujeto u objeto i en la característica j, o al rango asignado al sujeto u objeto i por el juez j; y R_i se refiere a la suma de los rangos correspondientes al sujeto u objeto i:

$$R_i = \sum_{j=1}^{J} R_{ij}$$

Podemos decir que se da concordancia perfecta entre J conjuntos de rangos cuando todos los jueces valoran o clasifican a los n sujetos u objetos del mismo modo (es decir, cuando los jueces coinciden plenamente en sus juicios) o cuando los n sujetos u objetos son clasificados de idéntica manera en las J características consideradas. Cuando esto ocurre, todos los jueces coinciden en asignar el rango 1 a uno de los sujetos u objetos, el rango 2 a otro de los sujetos u objetos, ..., el rango n a otro de los sujetos u objetos. Esto significa que los totales R_i correspondientes a los diferentes sujetos u objetos serán: $1J$, $2J$, $3J$, ..., iJ, ..., nJ.

Decimos, por el contrario, que no existe concordancia entre J conjuntos de rangos cuando los n sujetos u objetos son valorados o clasificados de diferente forma por los J jueces (es decir, cuando los jueces no coinciden en sus juicios) o cuando los n sujetos u objetos son clasificados de diferente manera en las J características consideradas. Cuando esto ocurre, a uno de los sujetos u objetos le corresponden rangos de 1 a n, a otro de los sujetos u objetos le corresponden igualmente rangos de 1 a n, y lo mismo con el resto de los sujetos u objetos. Lo cual implica que, en el caso de concordancia nula, los totales R_i correspondientes a los diferentes sujetos u objetos serán iguales:

$$R_1 = R_2 = \cdots = R_i = \cdots = R_n = \frac{J(n+1)}{2}$$

(pues la suma de los J conjuntos de rangos vale $Jn(n{+}1)/2$). Así pues, el grado de concordancia existente queda reflejado en la variabilidad entre los totales R_i de los diferentes sujetos u objetos. Cuando la concordancia entre J conjuntos de rangos es perfecta, la variabilidad entre los R_i es máxima; cuando la concordancia es nula, la variabilidad entre los R_i es mínima.
Teniendo esto en cuenta, podemos definir el estadístico:

$$S = \sum_{i=1}^{n} \left(R_i - \frac{J(n+1)}{2} \right)^2$$

el cual representa la variabilidad observada entre cada total R_i y el total que cabría esperar si la concordancia fuera nula. S valdrá cero cuando la concordancia existente sea nula (pues, en ese caso, todos los totales R_i serán iguales entre sí e iguales a $J(n+1)/2$) y alcanzará su valor máximo en el caso de concordancia perfecta, es decir, cuando entre los totales R_i exista la máxima variabilidad:

$$S_{máx} = \frac{J^2 n(n^2 - 1)}{12}$$

Ahora bien, si queremos obtener un coeficiente que valga 0 en el caso de concordancia nula y 1 en el caso de concordancia perfecta* podemos servirnos de una transformación consistente en dividir S entre su valor máximo posible. Esta solución es justamente lo que conocemos como coeficiente de concordancia $\hat{W}$ de Kendall**:

$$\hat{W} = \frac{12 \sum_i R_i^2}{J^2 n(n^2 - 1)} - \frac{3(n+1)}{n-1}$$

Cuando entre J conjuntos de rangos existe concordancia máxima, $\hat{W}$ vale 1; cuando se da concordancia nula, $\hat{W}$ vale 0.

Para poder afirmar que existe concordancia significativa entre J conjuntos de rangos necesitamos hacer inferencias sobre el parámetro W. Esto, en realidad, tiene fácil solución pues $\hat{W}$ es fácilmente transformable en el estadístico X_r^2 de Friedman (ver apartado anterior):

$$X_r^2 = J(n-1)\,\hat{W}$$

De hecho, el coeficiente $\hat{W}$ de Kendall y el estadístico X_r^2 de Friedman son aplicables al mismo tipo de situaciones. Mantener la hipótesis de que las distribuciones poblacionales son idénticas dentro de cada sujeto o bloque utilizando el estadístico de Friedman es exactamente la misma cosa que mantener mediante el coeficiente de concordancia de Kendall la hipótesis de que las sumas de los J rangos asignados a cada sujeto u objeto (los totales $_{Ir}$) son iguales. Es decir, mantener la hipótesis nula de que los tratamientos son iguales es exactamente lo mismo que mantener la hipótesis nula de ausencia de concordancia.

---

* Con J conjuntos de rangos no tiene sentido un coeficiente con valores negativos, pues no es posible encontrar un desacuerdo total. Si entre dos conjuntos de rangos existe relación perfecta negativa, el tercer conjunto de rangos necesariamente estará relacionado con uno de los dos anteriores o con ninguno de ellos; y lo mismo vale decir del cuarto, y del quinto, etc.; y eso es algo de lo que no tiene sentido hablar en términos negativos.

** La presencia de empates dentro de un mismo conjunto de rangos hace que $\hat{W}$ tome un valor más pequeño del que le corresponde. El SPSS utiliza el coeficiente de Kendall aplicando una corrección por empates:

$$\hat{W} = \frac{12 \sum_i R_i^2 - 3J^2 n(n+1)^2}{J^2 n(n^2 - 1) - J \sum_1^k (t_i^3 - t_i)}$$

donde t_i se refiere al número de puntuaciones empatadas en un rango dado y k al número de rangos distintos en los que se produce empate.

Prueba de Cochran

Al estudiar más de dos proporciones relacionadas nos encontramos en una situación similar a la expuesta para el caso de dos proporciones relacionadas. Seguimos trabajando con una variable que sólo puede tomar dos valores (variable dicotómica o dicotomizada), sólo que ahora tenemos más de dos ($J > 2$) muestras relacionadas.

El diseño es bastante simple: a n sujetos se le toman J medidas de una variable dicotómica, o J variables dicotómicas son medidas en una muestra de n sujetos. Estamos, por tanto, ante un diseño idéntico al presentado a propósito del ANOVA A-EF-MR (medidas repetidas o bloques con un sujeto por nivel y bloque), pero con la diferencia de que, aquí, la variable medida (es decir, la variable dependiente) es una variable dicotómica (una variable que sólo puede tomar dos valores).

Las proporciones marginales P_{+j} representan las proporciones de *aciertos* de cada muestra o tratamiento: $P_{+j} = T_{+j}/n$ (siendo T_{+j} la suma de *aciertos* de cada muestra). Si las J muestras proceden de poblaciones idénticas, cabe esperar que las proporciones marginales P_{+j} sean iguales, excepto en la parte atribuible a las fluctuaciones propias del azar muestral. Basándose en este hecho, Cochran (1950) ha diseñado un procedimiento* que permite poner a prueba la hipótesis de igualdad entre J proporciones poblacionales (H_0: $\pi_{+1} = \pi_{+2} = \cdots = \pi_{+J}$):

$$Q = \frac{J(J-1)\sum T_{+j}^2 - (J-1)T^2}{JT - \sum T_{i+}^2}$$

El estadístico Q de Cochran se distribuye según χ^2 con J–1 grados de libertad. El SPSS ofrece como nivel crítico la probabilidad de obtener valores Q iguales o mayores que el encontrado.

Ejemplo (Pruebas no paramétricas > Varias muestras relacionadas)

Este ejemplo muestra cómo obtener e interpretar las pruebas incluidas en el procedimiento **Pruebas no paramétricas** > **K muestras relacionadas**. Comenzaremos con la prueba de Friedman y el coeficiente de concordancia W (que permiten analizar datos ordinales) y terminaremos con la prueba de Cochran (que permite analizar datos nominales).

Prueba de Friedman y coeficiente de concordancia *W* de Kendall

Vamos a comenzar utilizando la prueba de *Friedman* y el *coeficiente de concordancia W de Kendall* para analizar los datos ya estudiados en el capítulo 16 a propósito del ANOVA de un factor con medidas repetidas. Comenzamos reproduciendo en la tabla 19.21 los datos que vamos a analizar.

---

* Este procedimiento es generalización del de McNemar para dos proporciones relacionadas. De hecho, si $J = 2$, el estadístico de McNemar y el de Cochran son exactamente el mismo (ver, por ejemplo, Conover, 1980, pág. 204).

Tabla 19.21. Datos de un diseño de un factor con medidas repetidas.

Sujetos	hora	día	semana	mes
1	16	8	8	12
2	12	9	9	10
3	12	10	10	8
4	15	13	7	11
5	18	12	12	12
6	13	13	8	10
7	18	16	10	13
8	15	9	6	6
9	20	9	11	8

Recordemos que se trata de un experimento diseñado para estudiar el efecto del paso del tiempo sobre la calidad del recuerdo: a un grupo de 9 sujetos se les hace memorizar una historia durante 20 minutos; más tarde, al cabo de una hora, de un día, de una semana y de un mes, se les pide que intenten memorizar la historia escribiendo todo lo que recuerden; un grupo de expertos evalúa la calidad del recuerdo de cada sujeto hasta elaborar los datos que muestra la tabla 19.21.

Se trata de un diseño de un factor (al que podemos llamar *tiempo*) con cuatro niveles (los cuatro momentos en los que se registra el recuerdo: al cabo de una hora, de un día, de una semana y de un mes) y una variable dependiente (la calidad del recuerdo). Puesto que a todos los sujetos se les registra el recuerdo en los cuatro momentos definidos por la variable factor, se trata de un factor de medidas repetidas. Y ya sabemos que hablar de *medidas repetidas* es equivalente a hablar de *muestras relacionadas*.

En consecuencia, tanto la prueba de Friedman como el coeficiente de concordancia W son estadísticos apropiados para analizar esta situación. Sin embargo, las hipótesis que permiten contrastar, aunque equivalentes, son distintas. El estadístico de Friedman contrasta la hipótesis de que los J promedios comparados son iguales en la población; el coeficiente de concordancia W contrasta la hipótesis de concordancia nula, es decir, la hipótesis de que los J conjuntos de puntuaciones comparados son independientes entre sí.

Para obtener la prueba de *Friedman* y el *coeficiente de concordancia W de Kendall*:

- Reproducir en el *Editor de datos* los datos de la tabla 19.21.
- En el cuadro de diálogo *Pruebas para varias muestras relacionadas* (ver figura 19.11), seleccionar las variables *hora*, *día*, *semana* y *mes* y trasladarlas a la lista **Contrastar variables**.
- En el recuadro **Tipo de prueba**, marcar la opción **Friedman** y **W de Kendall**.
- Pulsar el botón **Estadísticos...** para acceder al subcuadro de diálogo *Varias muestras relacionadas: Estadísticos* (ver figura 19.12) y marcar las opciones **Descriptivos** y **Cuartiles**.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran las tablas 19.22 a la 19.25. La tabla 19.22 ofrece algunos estadísticos descriptivos para las dos variables seleccionadas: el número de casos válidos en ambas variables, la media, la desviación típica, el valor más pequeño, el más grande y los cuartiles. Y a tabla 19.23 recoge, para cada variable, los rangos medios resultantes del proceso de asignación de rangos.

Tabla 19.22. Estadísticos.

	N	Media	Desviación típica	Mínimo	Máximo	Percentiles 25	Percentiles 50 (Mediana)	Percentiles 75
Una hora	9	15,00	2,29	12	18	12,50	15,00	17,00
Un día	9	11,00	2,65	8	16	9,00	10,00	13,00
Una semana	9	9,00	1,94	6	12	7,50	9,00	10,50
Un mes	9	10,00	2,29	6	13	8,00	10,00	12,00

Tabla 19.23. Rangos.

	Rango promedio
Una hora	3,94
Un día	2,44
Una semana	1,67
Un mes	1,94

La tabla 19.24 recoge la ***prueba de Friedman***. Contiene el número de casos válidos, el estadístico *chi-cuadrado* de Friedman, sus grados de libertad (*gl*) y el nivel crítico (*Sig. asintót.*). Puesto que el valor del nivel crítico (0,000) es menor que 0,05, podemos rechazar la hipótesis de igualdad de promedios poblacionales y concluir que la calidad del recuerdo no es la misma en los cuatro momentos considerados (*hora*, *día*, *semana* y *mes*).

Tabla 19.24. Prueba de Friedman.

N	9
Chi-cuadrado	18,556
gl	3
Sig. asintót.	,000

Aunque existen procedimientos para efectuar comparaciones múltiples cuando el estadístico de Friedman resulta significativo, (ver, por ejemplo, Pardo y San Martín, 1998, pág. 447), para analizar con el SPSS qué variables difieren entre sí podemos utilizar la prueba de Wilcoxon para dos muestras relacionadas, pero acompañada de la corrección de Bonferroni para controlar la tasa de error (es decir, para controlar la probabilidad de cometer errores de tipo I). Con cuatro variables necesitamos hacer seis comparaciones dos a dos: 1-2, 1-3, 1-4, 2-3, 2-4 y 3-4. La aplicación de la corrección de Bonferroni al hacer comparaciones por pares nos llevará a basar nuestras decisiones en un nivel de significación de 0,05/6 = 0,0083. Es decir, consideraremos que los promedios de dos variables difieren significativamente cuando el nivel crítico obtenido sea menor de 0,083.

La tabla 19.25 contiene la información relacionada con el ***coeficiente de concordancia W de Kendall***. La tabla muestra el número de casos válidos, el valor del estadístico *W*, su transformación en *chi-cuadrado* (que adopta exactamente el mismo valor que el estadístico de Friedman), sus grados de libertad y el nivel crítico (*Sig. asintót.*). Puesto que el valor del nivel crítico (0,000) es menor que 0,05, podemos rechazar la hipótesis de concordancia nula y concluir que entre las puntuaciones de las cuatro variables estudiadas existe asociación significativa.

Tabla 19.25. Coeficiente de concordancia *W* de Kendall.

N	9
W de Kendall	,687
Chi-cuadrado	18,556
gl	3
Sig. asintót.	,000

Conviene en este momento recordar que, aunque la prueba de Friedman permite comparar los promedios de *J* variables ordinales y el estadístico *W* de Kendall permite contrastar la presencia de asociación entre *J* variables ordinales, lo cierto es que ambas cosas son equivalentes. Dado que los rangos se asignan independientemente para cada sujeto, sólo es posible encontrar asociación entre *J* conjuntos de rangos si existen al menos dos promedios que difieren entre sí, y viceversa. De hecho, según hemos visto, el estadístico *chi-cuadrado* de la prueba de Friedman y el del coeficiente de concordancia *W* son exactamente el mismo.

Prueba de Cochran

Para ilustrar la prueba de Cochran vamos a utilizar un ejemplo tomado de Pardo y San Martín (1998, págs. 507-508). Imaginemos que deseamos averiguar si 4 preguntas de una prueba de rendimiento poseen o no la misma dificultad. Para ello, hacemos que una muestra aleatoria de 10 sujetos responda a las 4 preguntas y registramos las respuestas dadas: 1 = acierto; 2 = error. La tabla 19.26 muestra estos resultados.

Tabla 19.26. Aciertos (1) y errores (0) obtenidos por una muestra de 10 sujetos al responder a 4 preguntas de una prueba de rendimiento.

Sujetos	Pregunta 1	Pregunta 2	Pregunta 3	Pregunta 4
1	1	1	1	1
2	1	0	0	0
3	1	0	1	0
4	1	1	1	0
5	0	0	1	0
6	0	1	1	0
7	1	0	1	1
8	0	0	1	0
9	1	1	1	0
10	0	1	1	0

Si consideramos la proporción de aciertos de cada pregunta como un indicador de su grado de dificultad, ¿podemos afirmar que las preguntas difieren en grado de dificultad?

Puesto que los 10 sujetos responden a las 4 preguntas, nos encontramos ante un diseño de medidas repetidas o, lo que es lo mismo, de muestras relacionadas. Y dado que la variable dependiente o *respuesta* es dicotómica (acierto-error), debemos centrar nuestra atención en el análisis de *proporciones*. En consecuencia, lo apropiado en una situación de este tipo será utilizar la prueba de Cochran para *J* proporciones relacionadas. Para obtener la prueba de Cochran:

- Reproducir en el *Editor de datos* los datos de la tabla 19.26.
- En el cuadro de diálogo *Pruebas para varias muestras relacionadas* (ver figura 19.11), seleccionar las variables correspondientes a las preguntas 1, 2 3 y 4 (cualquiera que sea el nombre que se les haya asignado) y trasladarlas a la lista **Contrastar variables**.
- En el recuadro **Tipo de prueba**, marcar la opción **Cochran**.

Aceptando estas elecciones, el *Visor* ofrece los resultados que muestran las tablas 19.27 y 19.28. La primera de ellas recoge el número (frecuencia) de respuestas de cada tipo observado en cada pregunta.

Tabla 19.27. Frecuencias.

	Valor	
	0	1
Pregunta 1	4	6
Pregunta 2	5	5
Pregunta 3	1	9
Pregunta 4	8	2

La tabla 19.28 ofrece el número de casos válidos (*N*), el estadístico de Cochran (*Q de Cochran*), sus grados de libertad (*gl*) y su nivel crítico asintótico (*Sig. asint.*). Puesto que el nivel crítico (0,019) es menor que 0,05, podemos rechazar la hipótesis de igualdad de proporciones y concluir que la proporción de aciertos (y, consecuentemente, el grado de dificultad) no es la misma en las 4 preguntas.

Tabla 19.28. Prueba de Cochran.

N	10
Q de Cochran	10,000
gl	3
Sig. asintót.	,019

Aunque existen procedimientos para efectuar comparaciones múltiples cuando el estadístico de Cochran resulta significativo, (ver, por ejemplo, Pardo y San Martín, 1998, págs. 508-510), para contrastar con el SPSS qué proporciones difieren entre sí podemos utilizar la prueba de McNemar para dos muestras relacionadas (ya estudiada en el capítulo 12), pero acompañada de la corrección de Bonferroni para controlar la tasa de error (es decir, para controlar la probabilidad de cometer errores de tipo I).

Al igual que ocurría en el ejemplo anterior referido a la prueba de Friedman, con cuatro proporciones necesitamos hacer seis comparaciones dos a dos: 1-2, 1-3, 1-4, 2-3, 2-4 y 3-4. La aplicación de la corrección de Bonferroni al hacer comparaciones por pares nos llevará a basar nuestras decisiones en un nivel de significación de 0,05/6 = 0,0083. Por tanto, al comparar cada pareja de proporciones relacionadas utilizando la prueba de McNemar, sólo consideraremos que dos proporciones difieren significativamente entre sí cuando el nivel crítico obtenido al compararlas sea menor de 0,083.