PRUEBAS PARA DOS MUESTRAS RELACIONADAS

Estos contrastes permiten comprobar si hay diferencias entre las distribuciones de dos poblaciones a partir de dos muestras dependientes o relacionadas; es decir, tales que cada elemento de una muestra está emparejado con un elemento de la otra, de tal forma que los componentes de cada pareja se parezcan entre sí lo más posible por lo que hace referencia a un conjunto de características que se consideran relevantes. Tambén es posible que cada elemento de una muestra actúe como su propio control.

Algunas de la pruebas que pueden realizarse con el programa SPSS son: la prueba de Wilcoxon, la de signos y la de McNemar.

PRUEBA DE SUMA DE RANGOS DE WILCOXON

Cuando se trata de variables medibles en por lo menos una escala ordinal y pueden suponerse poblaciones contínuas la prueba no paramétrica más potente es la de Wilcoxon.

La hipótesis nula del contraste postula que las muestras proceden de poblaciones con la misma distribución de probabilidad; la hipótesis alternativa establece que hay diferencias respecto a la tendencia central de las poblaciones y puede ser direccional o no.

El contraste se basa en el comportamiento de las diferencias entre las puntuaciones de los elementos de cada par asociado, teniendo en cuenta no sólo el signo, sino también la magnitud de la diferencia.

Sea $d_i = x_i - y_i$ la diferencia entre las puntuaciones de la pareja i-ésima; si alguna de estas diferencias es nula la pareja correspondiente se elimina del análisis, de forma que el tamaño de la muestra es n, el número de diferencias no nulas. A continuación se asignan rangos desde 1 hasta n atendiendo únicamente al valor absoluto de las di y se suman los rangos correspondientes a las diferencias positivas y a las diferencias negativas por separado. Si la hipótesis nula es cierta, X e Y tienen el mismo valor central y es de esperar que los rangos se distribuyan aleatoriamente entre las diferencias positivas y negativas y, por tanto, que ambas sumas de rangos sean aproximadamente iguales. El estadístico de prueba, T, es la menor de las dos sumas de rangos. Cuando n > 15 la distribución muestral de T bajo el supuesto de que H0 es cierta se aproxima a una normal de parámetros:

$$\mu_T = \frac{n(n+1)}{4} \qquad \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

El estadístico de prueba es el valor Z:

$$Z = \frac{T - \mu_T}{\sigma_T}$$

que se distribuye según una normal tipificada.

Para el nivel de significación deseado se rechazará la hipótesis nula si Z pertenece a la región crítica localizada en las dos colas o en una cola de la normal tipificada, según la naturaleza de la hipótesis alternativa.

PRUEBA DE SIGNOS

La prueba de los signos permite contrastar la hipótesis de que las respuestas a dos ''tratamientos'' pertenecen a poblaciones idénticas. Para la utilización de esta prueba se requiere únicamente que las poblaciones subyacentes sean contínuas y que las respuestas de cada par asociado estén medidas por lo menos en una escala ordinal.

La hipótesis nula puede expresarse como:

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

Siendo Xi la respuesta del elemento i-ésimo al primer ''tratamiento'' e Yi la respuesta del elemento i-ésimo al segundo ''tratamiento''.

La hipótesis alternativa puede ser direccional, cuando postula que X es estocásticamente mayor (o menor) que Y, o no direccional, cuando no predice la dirección de la diferencia.

Para realizar el contraste se hallan los signos (+ o -) de las diferencias no nulas entre las respuestas de los dos componentes de cada par y se cuenta cuántas son positivas, S+, y cuántas negativas, S-. Si H0 es cierta, es de esperar que aproximadamente la mitad de las diferencias sean positivas y la otra mitad negativas.

El estadístico de prueba es *S= mín* [S+, S-].

Si H0 es cierta, S tiene distribución binomial de parámetros n= nº de diferencias nulas y π = 0'5. Si n es grande, la distribución de S puede aproximarse mediante una normal de parámetros $\mu_S = n\pi = 0,5n$ y $\sigma_S^2 = n\pi(1-\pi) = 0,25n$, y la decisión dependerá del valor tipificado de S. Para mejorar la aproximación se realiza una corrección de continuidad, de forma que el estadístico de prueba es:

$$Z = \frac{(S-0,5)-0,5n}{0,5\sqrt{n}}$$

Z se distribuye según una normal tipificada.

Cuando el número de diferencias no nulas es pequeño la aproximación de la distribución de S mediante la normal no es buena y en este caso el SPSS realiza directamente la prueba binomial, dando el nivel de significación a partir del cual se rechaza H0 en un contraste de dos colas. Si el contraste se realiza a una cola dicho nivel de significación se reduce a la mitad.

PRUEBA DE MCNEMAR

La prueba de McNemar se utiliza para decidir si puede o no aceptarse que determinado ''tratamiento'' induce un cambio en la respuesta dicotómica o dicotomizada de los elementos sometidos al mismo, y es aplicable a los diseños del tipo ''antes-después'' en los que cada elemento actúa como su propio control.

Los resultados correspondientes a una muestra de n elementos se disponen en una tabla de frecuencias 2 x 2 para recoger el conjunto de las respuestas de los mismos elementos antes y después. El aspecto general de dicha tabla, en la que los signos + y - se utilizan para representar las diferentes respuestas, es el siguiente:

Antes/Después	-	+
-	a	b
+	c	d

En las celdas de la tabla, *a* es el número de elementos cuya respuesta es la misma, -; *b* es el número de elementos cuya respuesta es - antes del ''tratamiento'' y + después de éste; *c* es el número de elementos que han cambiado de + a -; y *d* es el número de elementos que mantienen la respuesta +.

Por tanto, b+c es el número total de elementos cuyas respuestas han cambiado, y son los únicos que intervienen en el contraste. La hipótesis nula es que el ''tratamiento'' no induce cambios significativos en las respuestas, es decir, los cambios observados en la muestra se deben al azar, de forma que es igualmente probable un cambio de + a - que un cambio de - a +. Así pues, si H0 es cierta, de los b+c elementos cuya respuesta ha cambiado es de esperar que (b+c)/2 hayan pasado de + a -, y (b+c)/2 hayan pasado de - a +. En otras palabras, si H0 es cierta, la frecuencia esperada en las correspondientes celdas es (a+b)/2.

La hipótesis alternativa puede ser no direccional, cuando postula que la probabilidad de un cambio de + a - tiene distinta probabilidad que un cambio de - a +, o direccional, cuando predice que un cambio de - a + es más (o menos) probable que un cambio de + a -.

El estadístico de prueba que permite contrastar si existen diferencias significativas entre las frecuencias esperadas y las observadas es:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

O*i*= frecuencia observada en la i-ésima celda

E*i* = frecuencia esperada en la i-ésima celda si H0 es cierta

k = número de celdas

Para contrastar la significación de los cambios interesan sólo las celdas que recogen cambios, por tanto el estadístico puede expresarse como

$$\chi^2 = \frac{\left[b - \frac{b+c}{2}\right]^2}{\frac{b+c}{2}} + \frac{\left[c - \frac{b+c}{2}\right]^2}{\frac{b+c}{2}} = \frac{(b-c)^2}{b+c}$$

Si H0 es cierta, el estadístico χ^2 tiene distribución aproximadamente chi-cuadrado con 1 grado de libertad. La aproximación es más precisa si se realiza la corrección de continuidad de Yates, quedando el estadístico:

$$\chi^2 = \frac{(|b-c| - 1)^2}{b+c}$$

La hipótesis nula, de que ambos tipos de cambio son igualmente probables, se rechaza si el valor del estadístico se encuentra en la región crítica.

Cuando la frecuencia esperada (b+c)/2 es pequeña la aproximación de la distribución del estadístico de prueba a la chi-cuadrado no es buena y, en tal caso, el SPSS no calcula el estadístico anterior, sino que realiza la prueba binomial. El contraste se plantea en este caso de la siguiente forma:

supongamos que c<b; en este caso la hipótesis nula es que c es un valor de una variable X con distribución binomial de parámetros *n=b+c y* π *=0,5*. El nivel de significación para una prueba de dos colas es $P(X \leq c) + P(X \geq b)$y se rechazará H0 para niveles de significación iguales o superiores a éste. Si la hipótesis alternativa es direccional el nivel de significación a partir del cual se rechazará H0 es la mitad del nivel de significación bilateral.

REALIZACIóN DE LOS CONTRASTES

Para realizar estas pruebas la secuencia es:

Analizar

Pruebas no paramétricas

2 muestras relacionadas

En el cuadro de diálogo se seleccionan en *Contrastar pares* las variables. Se activa la o las pruebas que se quieren realizar en el recuadro *Tipo de prueba*. Por defecto únicamente está activada la prueba de *Wilcoxon*.

EJEMPLO

*Con los datos de la encuesta **Encinf.sav** probar si hay discrepancia entre la valoración que hacen los alumnos sobre la dotación de las aulas de informática (Dotacion) y la valoración que hacen del software disponible*

(Software).

Se trata de contrastar la hipótesis nula de que la valoración de la dotación de las aulas es igual a la valoración del software instalado. Dado que las valoraciones de ambas características son asignadas por los mismos individuos, las muestras resultantes no son independientes. Por otra parte, las variables se miden en una escala ordinal, y por tanto el contraste más adecuado es la prueba de Wilcoxon.

Para realizar este contraste la secuencia es:

Estadística > Pruebas no paramétricas > 2 muestras relacionadas.

En el cuadro de diálogo se selecciona en *Contrastar pares* las variables Dotacion y Software; por defecto está activada la prueba de *Wilcoxon*. Al aceptar se obtienen los siguientes resultados:

Rangos

		N	Rango promedio	Suma de rangos
SOFTWARE - dotación	Rangos negativos	19[a]	33,50	636,50
	Rangos positivos	66[b]	45,73	3018,50
	Empates	21[c]		
	Total	106		

a. SOFTWARE < dotación

b. SOFTWARE > dotación

c. dotación = SOFTWARE

Estadísticos de contraste[b]

	SOFTWARE - dotación
Z	-5,280[a]
Sig. asintót. (bilateral)	,000

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Como puede verse en el cuadro Rangos, el número de elementos para los cuales el valor de la variable Software es mayor que el de la variable dotacion es considerablemente mayor que el de los elementos para los que está mejor valorada la dotación que el software. En el cuadro Estadísticos de contraste, el valor tipificado del estadístico de prueba (la menor de las dos sumas de rangos) es igual a -5,280; por tanto, se rechaza la hipótesis nula de que la valoración de la dotación es igual que la del software para cualquier nivel de significación.

PRUEBAS PARA K MUESTRAS INDEPENDIENTES

En este apartado se presentan dos pruebas que permiten contrastar si k >2 muestras aleatorias e independientes proceden de una misma población, es decir, si un factor que subdivide la población de origen incide de forma significativa sobre el valor central de la población. Estos contrastes son alternativas no paramétricas al análisis de la varianza cuando se incumple alguno de los supuestos básicos de dicho análisis. El único requisito para aplicar estos contrastes es que la variable esté medida al menos en una escala ordinal.

PRUEBA H DE KRUSKAL-WALLIS

Este contraste permite decidir si puede aceptarse la hipótesis de que k muestras independientes proceden de la misma población o de poblaciones idénticas con la misma mediana. El único supuesto necesario es que las distribuciones subyacentes de las variables sean continuas y que éstas hayan sido medidas por lo menos en una escala ordinal.
Sean n1, n2 ... nk los tamaños de cada una de las muestras y n el total de observaciones. Para el cálculo del estadístico de prueba se ordenan las n observaciones de menor a mayor y se les asignan rangos desde 1 hasta n. A continuación se obtiene la suma de los rangos correspondientes a los elementos de cada muestra, Rj y se halla el rango promedio. Si la hipótesis nula es cierta, es de esperar que el rango promedio sea aproximadamente igual para las k muestras; cuando dichos promedios sean muy diferentes es un indicio de que H0 es falsa.

El estadístico de prueba es:

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

Si H0 es cierta y los tamaños muestrales son todos mayores que 5, el estadístico H se distribuye aproximadamente como chi-cuadrado con k-1 grados de libertad. La aproximación es tanto mejor cuanto mayor es el número de muestras y el tamaño de las mismas.
Cuando se producen empates, es decir, cuando varias observaciones de la misma o de distintas muestras son iguales y a todas se les asigna el mismo rango, es necesario dividir el valor de H por el siguiente factor de corrección:

$$1 - \frac{\sum_{j=1}^{g}(t_i^3 - t_i)}{n^3 - n}$$

En esta expresión g es el número de rangos que se repiten y ti es el número de veces que se repite el rango i-ésimo. El efecto del factor de corrección es elevar ligeramente el valor de H.

PRUEBA DE LA MEDIANA

Mediante esta prueba se contrasta la hipótesis nula de que k muestras independientes de tamaños n1, n2 ... nk proceden de la misma población o de poblaciones con medianas iguales. Para este contraste se requiere que la variable sea medible por lo menos en una escala ordinal y es particularmente útil cuando por alguna razón (como, por ejemplo, por haberse establecido puntos de corte durante el proceso de obtención de los datos) se sabe que las muestras no pueden contener observaciones extremas.
Para hallar el valor del estadístico de prueba se ordenan las n observaciones (n1 + n2 +... +nk) de menor a mayor y se determina el valor de la mediana común, Me.. A continuación, cada una de las observaciones originales se asigna a una de dos categorías: a la categoría 1 si es superior a la mediana común o a la categoría 2 si es inferior o igual.
El número de observaciones de cada grupo asignadas a cada categoría se recoge en una tabla de contingencia como la siguiente:

Muestra	1	2	...	k	Total
Superiores a Me	n_{11}	n_{12}	...	n_{1k}	$\sum_{j=1}^{k} n_{1j}$
Inferiores a Me Inferior o iguales a Me	n_{21}	n_{22}	...	n_{2k}	$\sum_{j=1}^{k} n_{2j}$
Total	$\sum_{i=1}^{2} n_{i1}$	$\sum_{i=1}^{2} n_{i2}$	...	$\sum_{i=1}^{2} n_{ik}$	n

El estadístico de prueba es:

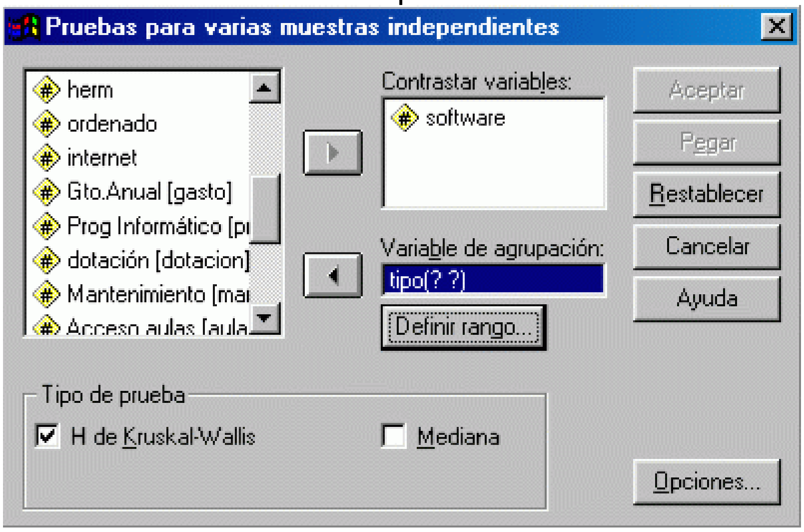
$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

Eij es la frecuencia esperada en la i-ésima fila de la j-ésima columna bajo el supuesto de que la hipótesis nula es cierta, calculada como producto de las frecuencias marginales dividido por n. Si la hipótesis nula es cierta el estadístico de prueba se distribuye aproximadamente como una chi-cuadrado con k - 1 grados de libertad. Se rechazará H0 si el valor del estadístico de prueba está en la región crítica.
Cuando aparecen frecuencias esperadas menores que 5 la aproximación de

la distribución del estadístico de prueba mediante la chi-cuadrado no es buena y deberán agruparse muestras o aumentar el tamaño de las mismas para resolver el problema.

REALIZACIóN DE LOS CONTRASTES

Para realizar estas pruebas la secuencia es:
Analizar
Pruebas no paramétricas
k muestras independientes



En el cuadro de diálogo se seleccionan en Contrastar variables la variable que recoge las puntuaciones de los grupos. En Variable de agrupación se indica la variable que determina los grupos, es decir, la variable que actúa como factor y se indica en Definir el rango los valores enteros para el máximo y el mínimo que se correspondan con las categorías mayor y menor de la variable de agrupación. Se activa la o las pruebas que se quieren realizar en el recuadro Tipo de prueba. Por defecto únicamente está activada la prueba H de Kruskal-Wallis.

EJEMPLO

Con los datos de la encuesta Enctrans.sav probar si los alumnos que utilizan habitualmente los transportes públicos (metro, bus, tren) valoran de forma significativamente distinta las características independencia (Inde) y rapidez (Rapi).

Se trata de contrastar la hipótesis nula de que la valoración asignada a la independencia y a la rapidez difieren significativamente en función del tipo de transporte público utilizado. Dado que las valoraciones de ambas características se miden en una escala ordinal y las muestras son independientes, el contraste más adecuado es la prueba H de Kruskal-Wallis.

Para realizar este contraste la secuencia es:

Estadística > Pruebas no paramétricas > k muestras independientes.

En el cuadro de diálogo se selecciona en Contrastar variables Independencia y Rapidez; en Variable de agrupación se indica el factor, es decir, la variable que induce los diferentes grupos, que en este caso es la variable Trans. Como únicamente interesa comparar la opinión de los usuarios del transporte público en el cuadro de diálogo que se abre con el botón Definir rango se indica como rango Mínimo 1 y como rango Máximo 3, ya que 1, 2 y 3 son las codificaciones asignadas a las modalidades metro, bus y tren respectivamente. Al aceptar se obtienen los siguientes resultados:

Rangos

	tipo de transporte	N	Rango promedio
rapidez	1 metro	53	52,49
	2 bus	29	37,34
	3 tren	13	53,46
	Total	95	
independencia	1 metro	46	43,70
	2 bus	25	38,26
	3 tren	12	43,29
	Total	83	

Estadísticos de contraste[a,b]

	rapidez	independencia
Chi-cuadrado	6,449	,891
gl	2	2
Sig. asintót.	,040	,640

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: tipo de transporte

Por lo que se refiere a la variable Rapidez, el estadístico de prueba es 6,449 y por tanto se rechaza la hipótesis nula según la cual los tres grupos valoran igualmente esta característica. En el caso de la variable Independencia el valor del estadístico Chi-cuadrado es 0,891 y no se rechaza la hipótesis nula

PRUEBAS PARA K MUESTRAS DEPENDIENTES

Cuando las k muestras están relacionadas de forma que las características de los i-ésimos elementos de cada muestra son idénticas o lo más parecidas posible, las diferencias observadas entre las muestras serán atribuidas únicamente al efecto del factor diferenciador de los grupos. El contraste de la hipótesis de que las k muestras proceden de una misma población o de poblaciones con la misma tendencia central no puede realizarse mediante el análisis de la varianza, al incumplirse el supuesto, por lo menos, de independencia de las muestras. En este caso puede utilizarse alguna de las alternativas no paramétricas que se presentan a continuación.

PRUEBA DE FRIEDMAN

Esta prueba puede utilizarse en aquellas situaciones en las que se seleccionan n grupos de k elementos de forma que los elementos de cada grupo sean lo más parecidos posible entre sí, y a cada uno de los elementos del grupo se le aplica uno de entre k ''tratamientos'', o bien cuando a cada uno de los elementos de una muestra de tamaño n se le aplican los k ''tratamientos''.

La hipótesis nula que se contrasta es que las respuestas asociadas a cada uno de los ''tratamientos'' tienen la misma distribución de probabilidad o distribuciones con la misma mediana, frente a la hipótesis alternativa de que por lo menos la distribución de una de las respuestas difiere de las demás. Para poder utilizar esta prueba las respuestas deben ser variables contínuas y estar medidas por lo menos en una escala ordinal.

Los datos se disponen en una tabla en la que en cada fila se recogen las respuestas de los k elementos de cada grupo a los k tratamientos:

Grupo\ Tratamiento	1	2	...	j	...	k
1	x_{11}	x_{12}	...	x_{1j}	...	x_{1k}
...	....	....	...	....	...	....
i	x_{i1}	x_{i2}	...	x_{ij}	...	x_{ik}
...	...	...	...	...	...	...
n	x_{n1}	x_{n2}	...	x_{nj}	...	x_{nk}

A las observaciones de cada fila se les asignan rangos de menor a mayor desde 1 hasta k; a continuación se suman los rangos correspondientes a cada columna, siendo R_J la suma correspondiente a la columna j-ésima. Si la hipótesis nula es cierta, la distribución de los rangos en cada fila se debe al azar, y es de esperar que la suma de los rangos correspondientes a cada columna sea aproximadamente igual a n(k + 1)/2. La prueba de Friedman determina si las R_J observadas difieren significativamente del valor esperado bajo la hipótesis nula.

El estadístico de prueba es:

$$F = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)$$

Si H0 es cierta y el número de columnas y/o de filas es moderadamente grande la distribución de F se aproxima a una chi-cuadrado con k - 1 grados de libertad; de forma que se rechaza la hipótesis nula para valores de F superiores al valor crítico para el nivel de significación fijado.

PRUEBA Q DE COCHRAN

Cuando sobre n elementos se observa la serie de respuestas de cada uno de ellos a k ''tratamientos'' esta prueba permite contrastar la hipótesis nula de que no existe diferencia significativa entre los k ''tratamientos''. También es posible utilizarla si cada tratamiento se aplica a uno de los elementos de n grupos de k elementos elegidos de forma que los elementos de cada grupo se asemejen lo más posible entre ellos.

Esta prueba es adecuada cuando la respuesta a cada tratamiento es una variable dicotómica, siendo X = 1 si la respuesta es ''éxito'' y X = 0 si es ''no éxito'' Si la respuesta es susceptible de medición en por lo menos una escala ordinal también es posible dicotomizarla, pero se pierde información y, por lo tanto, es preferible utilizar la prueba de Friedman.

Los datos se disponen en una tabla de la misma forma que para la prueba de Friedman, pero ahora las columnas de la tabla contienen únicamente ceros y unos, de forma que la suma de los valores de la j-ésima columna, G_J , es el número de ''éxitos'' de la distribución de las n respuestas al j-ésimo ''tratamiento''. Si la hipótesis nula es cierta las diferencias entre el número de éxitos de cada columna se deben al azar, por lo que es de esperar que sean pequeñas, es decir, que todas las G_{j estén muy próximas al número medio de éxitos por muestra, $\overline{G}.$El estadístico de prueba se basa en la dispersión del número de éxitos de cada "tratamiento" con respecto a $\overline{G}.$:

El estadístico de prueba es:

$$Q = \frac{k(k-1)\sum_{j=1}^{k}(G_j - \overline{G})^2}{k\sum_{i=1}^{n} L_i - \sum_{i=1}^{n} L_i^2}$$

Donde L*i* es el total de "éxitos" del primer elemento o grupo.

Si la hipótesis nula es cierta, $n \geq 4$ y $nk \geq 24$ la distribución de *Q* puede aproximarse mediante una chi-cuadrado con k - 1 grados de libertad y se rechaza la hipótesis nula si el valor de *Q* es superior al valor crítico para el nivel de significación deseado.

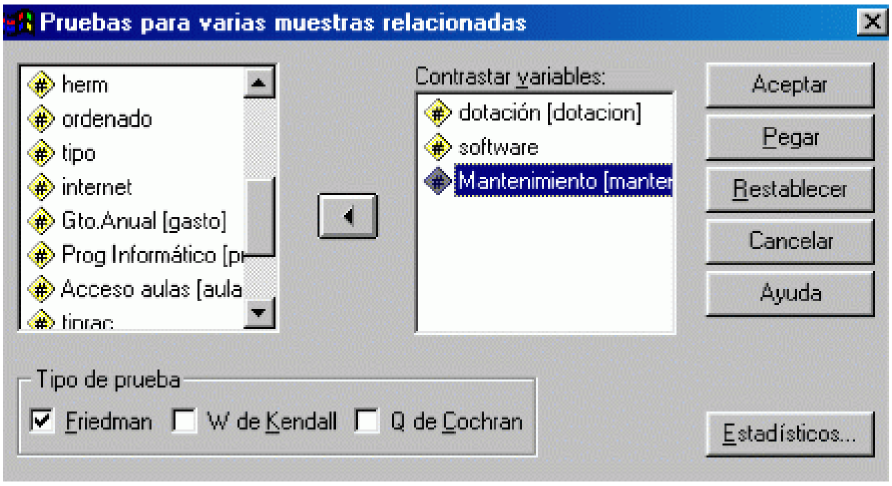
REALIZACIóN DE LOS CONTRASTES

Para realizar estas pruebas la secuencia es:

Analizar

Pruebas no paramétricas

k muestras relacionadas



En el cuadro de diálogo Pruebas para varias muestras relacionadas se indican las variables que se quiere comparar y se activan las pruebas que se desea realizar. Por defecto está activada la prueba de Friedman.

EJEMPLO

Con los datos de la encuesta ***Encinf.sav*** *probar si hay discrepancia entre la valoración que hacen los alumnos al mantenimiento (Manten), acceso a las aulas de informática (Aulas) y la valoración que hacen a los monitores que supervisan las aulas (Monitor).*

Como se trata de contrastar la hipótesis nula de que las valoraciones asignadas por los alumnos a las características mantenimiento, acceso y monitores de las aulas no difiere significativamente a partir de las puntuaciones asignadas por los mismos individuos, las muestras resultantes no son independientes. Por otra parte, las variables se miden en una escala ordinal, y por tanto el contraste más adecuado es la prueba de Friedman.

Para realizar este contraste la secuencia es:

Analizar > Pruebas no paramétricas > k muestras relacionadas.

En el cuadro de diálogo se seleccionan las variables Manten, Aulas y Monitor y se mantiene el tipo de prueba activado por defecto, Friedman.

Los resultados que se obtienen son los siguientes:

Rangos

	Rango promedio
Mantenimiento	2,20
Acceso aulas	1,95
monitores de aula	1,85

Estadísticos de contraste[a]

N	104
Chi-cuadrado	8,040
gl	2
Sig. asintót.	,018

a. Prueba de Friedman

El estadístico de prueba es igual a 8,040, por lo que no se puede rechazar la hipótesis nula para niveles de significación superiores a 0,018. Al 5% de nivel de significación se acepta la hipótesis de que no existen diferencias significativas entre las valoraciones asignadas por los alumnos a estas características.