T. 5 – Pruebas no paramétricas

1. Los supuestos del contraste de hipótesis

• Los diferentes tipos de contraste de hipótesis revisados en los temas precedentes implican para su aplicación que se cumplan algunas condiciones. Estas condiciones o supuestos pueden variar de un contraste de hipótesis a otro, si bien, los dos siguientes afectan a una parte importante de ellos:

1. De los contrastes de hipótesis hasta ahora descritos, un supuesto común a todos aquellos en que aparece implicada una o más variables cuantitativas es que éstas se distribuyan normalmente en la población de referencia.

   El cumplimiento de este supuesto puede ser comprobado a través de la prueba de Kolmogorov-Smirnov o la de Shapiro-Wilk, no obstante, diversos estudios han puesto de manifiesto que la importancia de este supuesto se relativiza a medida que el tamaño muestral es mayor, en especial a partir de muestras mayores de 30 casos.
2. Otro supuesto que también afecta a aquellos contrastes en que participan una o más variables cuantitativas es el que hace referencia a la propia naturaleza de estas variables,

esto es, que el nivel de medida de las mismas sea cuantitativo, no cuasi-cuantitativo o categórico. Esta es una condición que ha generada no poca controversia en la investigación en las Ciencias Sociales y de la Salud, pues es discutible si la escala de medida utilizada para medir algunas variables da lugar a una variable cuantitativa o a una ordinal – es el caso de muchos tests y escalas psicológicas.

3. Una doble condición común a los contrastes de hipótesis hasta ahora descritos en que aparecía implicada una variable categórica y una cuantitativa es la siguiente:
   3.1. Que la variable cuantitativa tenga igual varianza en las diferentes sub-poblaciones definidas por la variable categórica. Éste es el conocido como supuesto de homogeneidad de las varianzas y, como ya se expuso en un capítulo previo, su cumplimiento puede ser contrastado con la prueba de Levene.
   3.2. Que la variable cuantitativa se distribuya normalmente en las diferentes sub-poblaciones definidas por la variable categórica. La importancia de que se cumpla este supuesto es menor si se trabaja con muestras grandes.

• Los procedimientos que se describen en este capítulo, más conocidos en la literatura estadística como tests o <u>pruebas no paramétricas</u>, son pruebas de significación que no presuponen el cumplimiento de los anteriores supuestos, por lo que debería considerarse su aplicación en aquellos casos en que alguno de los supuestos enumerados en el párrafo anterior se encuentre comprometido. Ahora bien, no hay que olvidar que estas pruebas no están exentas del cumplimiento de un par de supuestos que, por lo esencial de los mismos, son comunes a cualquier prueba de significación:

1. Que la muestra sea representativa de la población objeto de estudio.
2. Que las observaciones sean independientes, esto es, que los datos de cualquier caso en las variables medidas no estén condicionados por los datos de otros casos en la muestra.

• Frente a la ventaja de ser menos exigentes a nivel de supuestos, un <u>aspecto negativo</u> de la aplicación del contraste de hipótesis basado en pruebas de significación no paramétricas es que suelen ofrecer menor potencia estadística ($1-\beta$) que sus correspondientes ‘pruebas de significación paramétricas’, esto es, con las pruebas no paramétricas es menor la probabilidad de rechazar la H_o siendo verdadera la H_e.

• En la presentación de las pruebas no paramétricas que sigue nos hemos limitado a concretar qué tipo de contraste de hipótesis se puede llevar a cabo con las mismas, así como aspectos prácticos asociados a su aplicación, pero no a detallar como se realiza el cálculo de los respectivos estadísticos de contraste, el cual se asume que podrá llevarse a cabo con un paquete estadístico –los ejemplos que se incluyen aquí han sido ejecutados con el programa *SPSS*.

2. Contraste de hipótesis sobre una variable: la prueba de los signos

• Tanto la prueba de Wilcoxon como la prueba de los signos permiten el contraste de hipótesis acerca de la localización de una variable, pero no respecto a la media (como la prueba *T* para una muestra) sino respecto a la mediana. Así, la hipótesis estadística para ambas pruebas, para el caso concreto de un contraste bilateral es la siguiente:

$$H_e : Mdn_X \neq k \rightarrow H_o : Mdn_X = k$$

• La prueba de Wilcoxon requiere que la variable a analizar sea cuantitativa y que la distribución de la misma sea simétrica, supuestos que no lo son para la prueba de los signos, que tan sólo asume que la variable a analizar sea al menos ordinal. Debido a esta mayor simplicidad en sus supuestos, nos centramos a continuación en la prueba de los signos, también referida en la literatura estadística como prueba binomial.

Ejemplo: Se deseaba contrastar si el salario medio en una gran empresa era superior a 1400€, para lo que se decidió a partir de los datos de una muestra aleatoria de 40 de sus miembros aplicar la prueba de los signos. Se optó por esta prueba de significación, en lugar de la prueba *T* para una muestra, debido el carácter habitualmente asimétrico positivo de la variable "Salario" –resultaría difícilmente justificable asumir la normalidad de esta variable para la población de una gran empresa. Así, las hipótesis para este contraste serían:

$$H_e : Mdn_X > 1400 \quad \text{y, complementariamente,} \quad H_o : Mdn_X \leq 1400$$

A continuación se muestran algunos estadísticos descriptivos de la citada variable para nuestra muestra, donde se ponen de manifiesto dos aspectos relevantes: por una parte, el carácter no normal de la variable "Salario" para los datos de nuestra muestra, lo cual alienta nuestras sospechas de no normalidad para esta variable; por otra parte, que la mediana es igual a 1450, lo cual en principio ofrece evidencia a favor de la H_e y, en consecuencia, da pie a que sigamos en la realización del contraste de hipótesis.

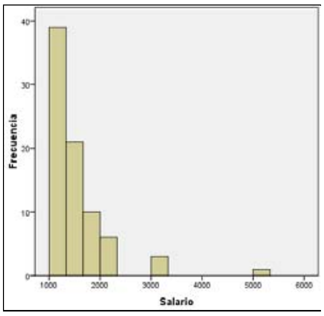
Salario

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1100	10	12,5	12,5	12,5
	1300	29	36,3	36,3	48,8
	1450	21	26,3	26,3	75,0
	1800	10	12,5	12,5	87,5
	2200	6	7,5	7,5	95,0
	3000	3	3,8	3,8	98,8
	5000	1	1,3	1,3	100,0
	Total	80	100,0	100,0	

Estadísticos

Salario

N	Válidos	80
	Perdidos	0
Media		1554,38
Mediana		1450,00
Moda		1300



Mostramos ahora el camino a seguir con *SPSS* en la realización de la prueba de los signos o binomial, así como la salida de resultados de este programa para los datos de nuestro ejemplo:

SPSS: Analizar > Pruebas no paramétricas > Binomial (Punto de corte = *k*)

Prueba binomial

		Categoría	N	Proporción observada	Prop. de prueba	Sig. asintót. (bilateral)
Salario	Grupo 1	<= 1400	39	,49	,50	,911[a]
	Grupo 2	> 1400	41	,51		
	Total		80	1,00		

a. Basado en la aproximación Z.

Si nos fijamos en el nivel de significación, éste es igual a 0,91, si bien, este valor es para un contraste bilateral, por lo que debemos dividirlo entre 2 por tratarse nuestro ejemplo de un contraste unilateral: 0,911/2 = 0,455. Suponiendo que hubiésemos asumido un nivel de riesgo igual a 0,05 en el contraste de hipótesis, el nivel de significación obtenido (P = 0,455) no nos permite rechazar la H_o. Así, la evidencia no apoya que el salario medio en la población de referencia esté por encima de los 1400€.

Si hubiésemos asumido la normalidad de la variable "Salario", al realizar este mismo contraste de hipótesis a través de la prueba T para una muestra se obtienen los resultados que se muestran en la siguiente tabla de resultados de *SPSS*, los cuales sustentan el rechazo de la H_o (con un α del 0,05) y, por tanto, apoyarían la conjetura inicial de nuestro estudio, en este caso, $H_e : \mu_{Salario} > 1400$.

Prueba para una muestra

	Valor de prueba = 1400					
					95% Intervalo de confianza para la diferencia	
	t	gl	Sig. (bilateral)	Diferencia de medias	Inferior	Superior
Salario	2,434	79	,017	154,375	28,15	280,60

3. Contraste de hipótesis sobre la relación entre dos variables

• A continuación se describen una serie de pruebas no paramétricas orientadas a llevar a cabo el contraste de hipótesis sobre la asociación entre dos variables. Uno de los puntos fuertes de estas pruebas en la práctica es que pueden ser aplicadas cuando no se satisfagan los supuestos para algunas de las pruebas vistas en los capítulos precedentes. En concreto, se van a tratar los que podríamos denominar como *equivalentes no paramétricos* de la prueba *T* para muestras independientes, la prueba *T* para muestras relacionadas, el ANOVA de un factor entre-sujetos, el ANOVA de un factor intra-sujetos y el coeficiente de correlación de Pearson.

La siguiente tabla ofrece un clasificación que nos puede ayudar a la hora de discriminar las condiciones de aplicación de las cuatro pruebas no paramétricas que van a ser tratadas primero a continuación.

	Muestras independientes	**Muestras relacionadas**
Variable categórica dicotómica	*Prueba de Mann-Whitney*	*Prueba de los signos para dos muestras*
Variable categórica politómica	*Prueba de Kruskal-Wallis*	*Prueba de Friedman*

3.1. La prueba de Mann-Whitney

• Esta prueba no paramétrica permite contrastar si es estadísticamente significativa la relación entre una variable categórica dicotómica y una variable cuantitativa (u ordinal), haciéndose operativo este contraste a través de la comparación de una estimación basada en valores de orden (también denominados, rangos) de la posición de los dos subgrupos de casos definidos por la variable categórica. Así, en el caso de un contraste bilateral, tendremos las siguientes hipótesis donde δ representa la diferencia en la localización de las dos poblaciones:

$$H_e : \delta \neq 0 \quad \text{y, complementariamente,} \quad H_o : \delta = 0$$

Ejemplo: En una replicación realizada del estudio de Loftus y Burns (ya presentado en un capítulo previo, recordamos que tenía por objeto comprobar en qué medida un choque emocional puede alterar el recuerdo) se midió la variable “Recuerdo” con una escala de medida que tan sólo diferenciaba las tres siguientes categorías de nivel de recuerdo: bueno, regular y malo. Dada la dificultad para asumir el carácter cuantitativo de esta variable, aunque sí ordinal, se decidió aplicar la prueba de Mann-Whitney para así contrastar si existía una diferencia estadísticamente significativa en el recuerdo de aquéllos que recibieron choque emocional y aquéllos que no y, en consecuencia, contrastar la existencia de relación entre las variables “Recuerdo” y “Choque emocional [Si/No]” ($H_e : \delta \neq 0$; $\alpha = 0{,}05$). Los resultados del análisis de los datos de este estudio, en el que participó una muestra de 24 sujetos (12 en cada condición experimental), se muestran en el siguiente “output” de *SPSS*:

SPSS: Analizar > Pruebas no paramétricas > Dos muestras independientes: U de Mann-Whitney (Opciones: Estadísticos descriptivos)

Rangos

	Choque_emocional	N	Rango promedio	Suma de rangos
Recuerdo	No	12	15,50	186,00
	Si	12	9,50	114,00
	Total	24		

Estadísticos de contraste[b]

	Recuerdo
U de Mann-Whitney	36,000
W de Wilcoxon	114,000
Z	-2,261
Sig. asintót. (bilateral)	,024
Sig. exacta [2*(Sig. unilateral)]	,039[a]

a. No corregidos para los empates.

b. Variable de agrupación: Choque_emocional

El estadístico de contraste de esta prueba de significación recibe el nombre de U de Mann-Whitney. Su valor para nuestro ejemplo es igual a 36, siendo el nivel de significación asociado al mismo el que aparece en la tabla como *Sig. asintót. (bilateral)*, esto es, 0,024. En el caso de que se hubiese planteado un contraste unilateral, deberíamos dividir este valor de significación entre dos.

El nivel de significación obtenido en nuestro ejemplo conduce al rechazo de la hipótesis nula, por lo que el estudio realizado aportaría evidencia a favor de la existencia de una relación estadísticamente significativa entre ambas variables. Ahora bien, a fin de interpretar el sentido de esta relación deberemos fijarnos en la primera tabla de resultados (la titulada *Rangos*), en concreto, en la columna *Rango promedio*, pues el subgrupo con un rango promedio más alto –en nuestro ejemplo, el grupo que no recibió choque emocional–, será el que también tenga una posición más elevada en la variable cuantitativa –en nuestro ejemplo, el recuerdo–, información que nos permitirá aportar una interpretación sustantiva a la relación obtenida.

3.2. La prueba de los signos para dos muestras

• Tanto la prueba de Wilcoxon para dos muestras como la prueba de los signos para dos muestras permiten el contraste de hipótesis acerca de la relación entre una variable categórica dicotómica y una variable cuantitativa/ordinal en que esta última es medida en una de las dos siguientes circunstancias: (1) en un mismo grupo de sujetos antes y después de la aplicación de una determinada acción (intervención, tratamiento...) que viene representada por la variable dicotómica (antes *versus* después); (2) en dos grupos de sujetos distintos pero relacionados entre sí, esto es, cada sujeto en uno de los grupos tiene un par en el otro grupo con el que tiene algún tipo de equivalencia en terceras variables –un caso paradigmático de diseño de investigación en que se da esta circunstancia es aquél en que los dos grupos están constituidos por pares de gemelos..

• Las dos pruebas citadas representan el *equivalente no paramétrico* de la prueba *T* para muestras relacionadas. Ahora bien, la prueba de Wilcoxon para dos muestras requiere que una de las variables a analizar sea cuantitativa y que la distribución de la misma sea simétrica, supuestos que no lo son para la prueba de los signos para dos muestras, que tan sólo asume que una de las variables a analizar sea al menos ordinal.

• A continuación nos vamos a centrar en la prueba de los signos para dos muestras por ser menos restrictiva en sus supuestos de aplicación, si bien, tanto ésta como la prueba de Wilcoxon para dos muestras comparten la realización del contraste de hipótesis basado, no en la media (como la prueba *T* para dos muestras relacionadas), sino en la mediana. Así, la hipótesis estadística en ambas pruebas, para el caso concreto de un contraste bilateral, es la siguiente:

$$H_e : Mdn_B \neq Mdn_A \;\rightarrow\; H_o : Mdn_B = Mdn_A$$

Ejemplo: En un estudio se deseaba comprobar si una determinada acción comunitaria tenía un efecto favorecedor del comportamiento altruista de la población infantil. Para ello, se formaron dos grupos, uno recibió la intervención mientras que el otro no, y tres meses después se llevó a cabo un proceso de observación del que se derivó, para cada niño/a, una de las cuatro siguientes categorías de medida: comportamiento altruista destacado (4); ídem. moderado (3); ídem. contenido (2); ídem. exiguo (1). Los dos grupos se formaron seleccionando 20 pares de niños cuyas condiciones familiares y de entorno social eran análogas. Por tratarse de grupos relacionados y por haber sido medido el comportamiento altruista con una escala de medida ordinal, se decidió aplicar la prueba de los signos para dos muestras a fin de contrastar la existencia de relación entre la puesta en marcha de la acción comunitaria y el nivel de comportamiento altruista de la población infantil.

Los resultados del contraste de hipótesis ($H_e : Mdn_{SI} \neq Mdn_{NO}$; α = 0,05) acerca de esta relación se muestran en la siguiente salida de *SPSS* (los nombres asignados a las dos variables fueron “NO_Intervención” y “SI_Intervención”:

SPSS: Analizar > Pruebas no paramétricas > Dos muestras relacionadas: Signos

Frecuencias

		N
SI_Intervencion - NO_Intervencion	Diferencias negativas[a]	1
	Diferencias positivas[b]	11
	Empates[c]	8
	Total	20

a. SI_Intervencion < NO_Intervencion

b. SI_Intervencion > NO_Intervencion

c. SI_Intervencion = NO_Intervencion

Estadísticos de contraste[b]

	SI_ Intervencion - NO_ Intervencion
Sig. exacta (bilateral)	,006[a]

a. Se ha usado la distribución binomial.

b. Prueba de los signos

Dado que el nivel de significación (0,006) es menor que el nivel de riesgo asumido para el contraste de hipótesis (0,05) se rechaza la hipótesis nula, por lo que el estudio realizado aportaría evidencia a favor de la existencia de una relación estadísticamente significativa entre la aplicación de la intervención y la variable comportamiento altruista. A fin de interpretar el sentido de esta relación deberemos fijarnos en la tabla de resultados superior (*Frecuencias*) y, dado que el número de casos en los que mejora su posición en la variable "SI_Intervención" respecto a su posición en la variable "NO_Intervención" es mucho mayor (11) que el caso contrario (1), resulta evidente cuál es el sentido de la relación, esto es, puntuaciones más altas de comportamiento altruista aparecen asociadas a los sujetos que recibieron la intervención, mientras que puntuaciones más bajas aparecen asociadas a la muestra que no fue objeto de intervención comunitaria.

3.3. La prueba de Kruskal-Wallis

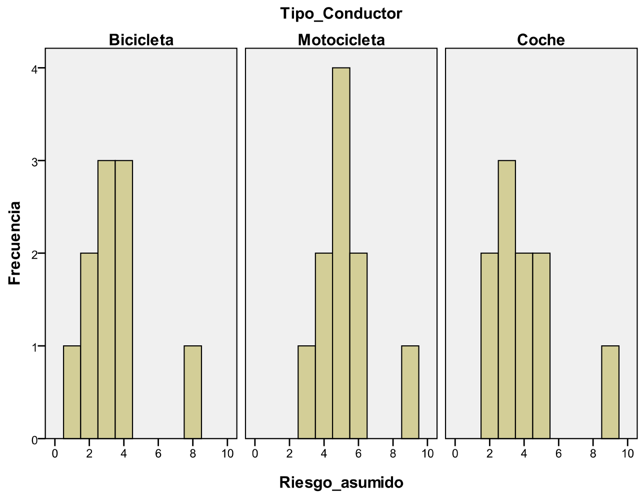
• Esta prueba no paramétrica permite contrastar si es estadísticamente significativa la relación entre una variable categórica y una variable cuantitativa/ordinal, independientemente del número de modalidades que tenga la variable categórica –en este sentido, puede considerarse como una generalización de la prueba de Mann-Withney para cualquier tipo de variable categórica.

• El contraste de hipótesis basado en la prueba de Kruskal-Wallis consiste en comparar una estimación basada en rangos de la posición de la variable cuantitativa/ordinal en los diferentes sub-muestras definidas por la variable categórica. Así, en el caso de un contraste bilateral, tendremos las siguientes hipótesis, donde δ representa la diferencia en la localización de las distintas sub-poblaciones objeto de comparación:

$$H_e : \delta \neq 0 \quad \text{y, complementariamente,} \quad H_o : \delta = 0$$

Ejemplo: Una investigación desarrollada por el INTRAS (Instituto de Investigación de Tráfico y Seguridad Vial de la Universitat de València) tuvo por objeto analizar la relación entre el tipo de conductor (bicicleta, motocicleta y coche) y el riesgo asumido en un conjunto de situaciones de tráfico reproducidas a través de un simulador de conducción. Se seleccionó a una muestra de 30 conductores, 10 que conducían habitualmente bicicleta, 10 moto y 10 coche. El riesgo asumido fue medido con una escala de 0 a 10, en la que una mayor puntuación representa un mayor riesgo asumido.

Diversos estudios han evidenciado que la variable "Riesgo asumido" suele distribuirse de forma asimétrica positiva. Los propios datos de este estudio ponen de manifiesto tal hecho, tal como puede observarse en el panel de histogramas obtenido para esta variable en cada uno de los tres subgrupos de conductores (ver más abajo). Es por ello que se decidió aplicar la prueba de Kruskal-Wallis, en vez del ANOVA para muestras independientes, a la hora de analizar si existían diferencias estadísticamente significativas (α = 0,05) en la posición de los tres tipos de conductores en el riesgo que asumen en la conducción.



El camino a seguir con el programa *SPSS* para realizar el contraste de hipótesis basado en la prueba de Kruskal-Wallis, así como los resultados de este análisis para los datos de nuestro ejemplo se muestran a continuación:

SPSS: Analizar > Pruebas no paramétricas > K muestras independientes: Kruskal-Wallis

Rangos

	Tipo_ Conductor	N	Rango promedio
Riesgo_asumido	Bicicleta	10	11,10
	Motocicleta	10	21,35
	Coche	10	14,05
	Total	30	

Estadísticos de contraste[a,b]

	Riesgo_ asumido
Chi-cuadrado	7,449
gl	2
Sig. asintót.	,024

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: Tipo_Conductor

El nivel de significación (*Sig* = 0,024) correspondiente al estadístico de contraste obtenido (*Chi-cuadrado* = 7,45) sustenta el rechazo de la H_o, proporcionando evidencia a favor de la existencia de una relación estadísticamente significativa entre el tipo de conductor (bicicleta, motocicleta y coche) y el riesgo asumido en la conducción. Además, la tabla de resultados (Rangos) permite explorar el sentido de esa relación, en concreto, como los conductores de motocicleta aparecen asociados a un mayor nivel riesgo asumido, seguidos a cierta distancia de los de coche y, en último lugar, los de bicicleta.

• Si de la prueba de Kruskal-Wallis se deriva la existencia de diferencias estadísticamente significativas, ello no significa que éstas se den entre todos los pares de subgrupos. Un análisis adicional basado en la prueba de Mann-Whitney nos permitirá conocer entre que pares concretos de subgrupos existen diferencias estadísticamente significativas, ahora bien, en el caso de llevar a cabo esa comparación por pares, será necesario realizar un ajuste en el nivel del α asumido (por ejemplo, la corrección de Bonferroni) a fin de contrarrestar el incremento en la tasa de error tipo I debido a estar realizando sobre unos mismos datos una serie de contrastes de hipótesis.

Si en nuestro **ejemplo** quisiéramos comparar por pares los tres tipos de conductores, tendríamos que realizar 3 comparaciones diferentes (bicicleta-motocicleta, bicicleta-coche y motocicleta-coche) por lo que, aplicando la corrección de Bonferroni (α dividido por el número de comparaciones), el α corregido a considerar en las 3 pruebas de Mann-Withney a realizar seria igual a 0,017 (0,05/3).

Así, al realizar las pruebas de Mann-Withney en nuestro ejemplo a fin de comparar por pares los tres tipos de conductores se obtuvieron los siguientes niveles de significación: bicicleta-motocicleta, 0,010; bicicleta-coche, 0,416; motocicleta-coche, 0,053. El nivel de significación es inferior a 0,017 solamente para la comparación entre los conductores de bicicleta y los de motocicleta, por lo que tan sólo para estos dos subgrupos se puede decir que existen diferencias estadísticamente significativas, esto es, diferencias en el riesgo asumido en la conducción a nivel poblacional.

3.4. La prueba de Friedman

• La prueba de Friedman puede considerarse una generalización de la prueba de los signos para dos muestras en cuanto que permite el contraste de hipótesis acerca de la relación entre una variable categórica y una variable cuantitativa/ordinal sin la restricción de que la variable categórica sea dicotómica. De modo análogo al ANOVA de un factor intra-sujetos, la prueba de Friedman se ajusta a diseños de recogida de datos en que la variable cuantitativa/ordinal es medida en una de las dos siguientes circunstancias: (1) en un mismo grupo de sujetos en diferentes momentos temporales (por ejemplo, antes de una intervención, un mes después de la intervención y 6 meses después de la intervención); (2) en dos o más grupos de sujetos relacionados entre sí, esto es, cada sujeto en uno de los grupos tiene sujetos parejos en los otros grupos respecto a terceras variables.

Ejemplo: Una investigación en el ámbito de un centro educativo tuvo por objeto analizar si una determinada intervención propuesta por el psicólogo escolar tenía efecto sobre el rendimiento educativo. Para ello, en una muestra de 12 estudiantes del grupo en que se llevo a cabo la intervención se midió su rendimiento educativo un mes antes de la intervención, un mes después y cinco meses después, para así valorar no sólo el efecto inmediato de la intervención, sino también su efecto a más largo plazo. El rendimiento educativo se midió con una escala (0-10) que se podía asumir como cuantitativa, por lo que la intención inicial fue aplicar el ANOVA intra-sujetos en el análisis de los datos, sin embargo, el incumplimiento de uno de los supuestos de este análisis condujo a la decisión de aplicar finalmente la prueba de Friedman.

Los pasos a seguir con *SPSS,* así como los resultados de la aplicación de la prueba de Friedman para los datos de este estudio se muestran a continuación:

SPSS: Analizar > Pruebas no paramétricas > K muestras relacionadas: Friedman (Opciones; Estadísticos descriptivos)

Estadísticos descriptivos

	N	Media	Desviación típica	Mínimo	Máximo
Antes_1mes	12	5,33	1,723	3	8
Despues_1mes	12	5,50	2,236	2	9
Despues_6mes	12	5,42	1,929	2	9

Rangos

	Rango promedio
Antes_1mes	1,83
Despues_1mes	2,17
Despues_6mes	2,00

Estadísticos de contraste[a]

N	12
Chi-cuadrado	,914
gl	2
Sig. asintót.	,633

a. Prueba de Friedman

El nivel de significación obtenido tras la aplicación de la prueba de Friedman (*Sig* = 0,633) sustenta que se mantenga la H_o, esto es, la no existencia de diferencias estadísticamente significativas entre las medidas del rendimiento académico en los tres momentos considerados. No se ha encontrado, por tanto, evidencia empírica a favor de la existencia de relación entre la intervención realizada y el cambio en el rendimiento académico en la población de referencia.

• Si la prueba de Friedman pone de manifiesto la existencia de diferencias estadísticamente significativas, un análisis adicional de comparación por pares basado en la prueba de los signos para dos muestras nos permitirá conocer entre que pares concretos de variables existen diferencias estadísticamente significativas.

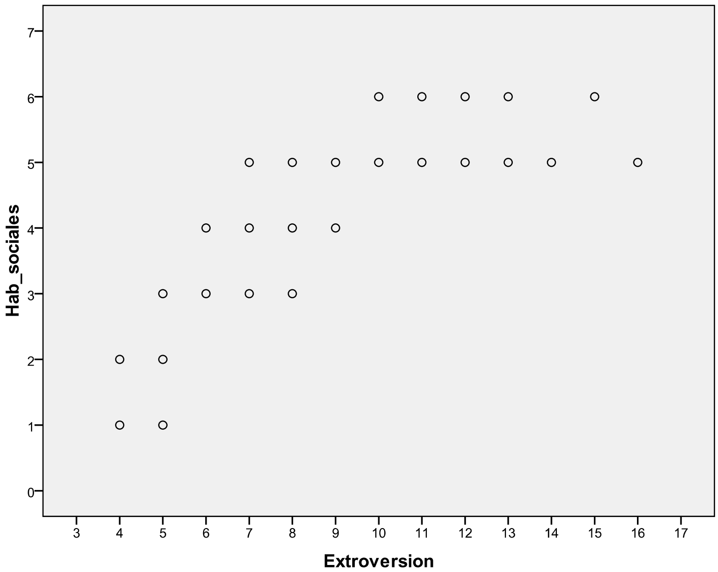
3.5. El coeficiente de correlación de Spearman

• Si no se cumplen los supuestos asociados a la aplicación del contraste de hipótesis del coeficiente de correlación de Pearson (esto es, que ambas variables se relacionen linealmente y se distribuyen normalmente en la población, siendo su nivel de medida cuantitativo), se puede optar por la utilización del coeficiente de correlación de Spearman (R_s) y su respectiva prueba de significación, cuyos supuestos menos restrictivos se limitan a que las variables sean al menos de tipo ordinal y que la relación entre ambas variables sea, al menos, monotónica, ya sea creciente (valores más altos en una variable aparecen asociados a valores más altos en la otra) o decreciente (valores más altos en una variable aparecen asociados a valores más bajos en la otra).

• La prueba de significación para el coeficiente de correlación de Spearman sirve para contrastar si existe relación entre dos variables cuantitativas/ordinales a nivel poblacional, por lo que la hipótesis nula del contraste representa la independencia entre ambas variables, o sea, que el valor de este coeficiente es igual a 0 a nivel poblacional ($\rho_S = 0$). Así, para el caso de un contraste bilateral:

$$H_e : \rho_S \neq 0 \text{ y, complementariamente, } H_o : \rho_S = 0$$

Ejemplo: El diagrama de dispersión que se presenta a continuación muestra la relación entre las puntuaciones en un test de extraversión (puntuaciones de 0 a 20) y las puntuaciones en habilidades sociales derivadas de la aplicación de un protocolo de observación (puntuaciones de 0 a 10) en una muestra 26 adolescentes que estaban cursando el bachillerato. El diagrama pone de manifiesto una relación monotónica creciente pero, dado lo cuestionable de asumir una relación lineal entre ambas variables a nivel poblacional, se decidió aplicar la prueba de significación para R_s a fin de realizar el contraste de hipótesis ($\alpha = 0{,}05$) acerca de la existencia de una relación estadísticamente significativa entre ambas variables.



Mostramos a continuación el camino a seguir con *SPSS* en la realización de la prueba de significación para R_s, así como la salida de resultados de este programa para los datos de nuestro ejemplo:

SPSS: Analizar > Correlaciones > Bivariadas: Spearman

Correlaciones

			Extroversion	Hab_sociales
Rho de Spearman	Extroversion	Coeficiente de correlación	1,000	,837**
		Sig. (bilateral)	.	,000
		N	26	26
	Hab_sociales	Coeficiente de correlación	,837**	1,000
		Sig. (bilateral)	,000	.
		N	26	26

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

• *SPSS* incluye los resultados de esta prueba de significación en la propia matriz de correlaciones de las variables analizadas, la cual en nuestro ejemplo sólo contiene la información de las dos variables objeto de análisis. El valor del coeficiente de correlación de Spearman (0,837) evidencia, por ser positivo, la existencia de una relación monótona creciente y, por lo próximo del mismo a 1, una relación alta entre las variables analizadas. En concordancia, el nivel de significación correspondiente a ese coeficiente es un valor muy bajo, inferior a 0,001, por lo que se rechazaría la

H_o y se concluiría que la evidencia apoya la existencia de una relación (monótona creciente) estadísticamente significativa entre las variables "Extroversión" y "Habilidades sociales".

Referencias

Losilla, J. M., Navarro, B., Palmer, A., Rodrigo, M. F., y Ato, M. (2005). *Del contraste de hipótesis al modelado estadístico.* Tarrasa: CBS (www.edicionsapeticio.com).

Loftus, E. F., y Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory and Cognition, 10,* 318-323.

Pardo, A., y San Martín, R. (2010). *Análisis de datos en ciencias sociales y de la salud II.* Madrid: Síntesis.