

<Artículo metodológico>

Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS.

Vanesa Berlanga Silvente y María José Rubio Hurtado

Fecha de presentación: 20/03/2012

Fecha de aceptación: 19/04/2012

Fecha de publicación: 04/07/2011

//Resumen

Las pruebas no paramétricas engloban una serie de pruebas estadísticas que tienen como denominador común la ausencia de asunciones acerca de la ley de probabilidad que sigue la población de la que ha sido extraída la muestra. Por esta razón es común referirse a ellas como pruebas de distribución libre. En el artículo se describen y trabajan las pruebas no paramétricas, y se resaltan su fundamento y las indicaciones para su empleo cuando se trata de una sola muestra (Chi-cuadrado), de dos muestras con datos independientes (U de Mann-Whitney), de dos muestras con datos relacionados (T de Wilcoxon), de varias muestras con datos independientes (H de Kruskal-Wallis) y de varias muestras con datos relacionados (Friedman).

//Palabras clave

Estadística no paramétrica, prueba no paramétrica, U de Mann-Whitney, T de Wilcoxon, H de Kruskal-Wallis, Friedman.

// Referencia recomendada

Berlanga Silvente, V. y Rubio Hurtado, M.J. (2012) Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. [En línea] *REIRE, Revista d'Innovació i Recerca en Educació*, Vol. 5, núm. 2, 101-113. Accesible en: <http://www.ub.edu/ice/reire.htm>

// Datos de las autoras

Vanesa Berlanga Silvente. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). berlanga.silvente@ub.edu

María José Rubio Hurtado. Profesora. Universidad de Barcelona. Departamento de Métodos de Investigación y Diagnóstico en Educación (MIDE). mjrubio@ub.edu



1. Introducción

En el ámbito de las Ciencias Sociales es habitual el uso de pruebas no paramétricas puesto que existen muchas variables que no siguen las condiciones de parametricidad. Dichas condiciones se refieren al uso de variables cuantitativas continuas, distribución normal de las muestras, varianzas similares y tamaño de las muestras, mayor a 30 casos. Estos criterios se recogen ampliamente en Rubio y Berlanga (2012). En caso de que no se cumplan estos requisitos, y sobre todo cuando la normalidad de las distribuciones de la variable en estudio esté en duda y el tamaño de la muestra sea menor a 30 casos, el empleo de las **pruebas no paramétricas o de distribución libre** está indicado.

Las pruebas no paramétricas reúnen las siguientes características: 1) son más fáciles de aplicar; 2) son aplicables a los datos jerarquizados; 3) se pueden usar cuando dos series de observaciones provienen de distintas poblaciones; 4) son la única alternativa cuando el tamaño de muestra es pequeño y 5) son útiles a un nivel de significancia previamente especificado.

2. Clasificación de las pruebas no paramétricas

La revisión de los principales autores que, en nuestro contexto, tratan la clasificación de las pruebas no paramétricas pone de manifiesto una falta de consenso a la hora de agrupar dichas pruebas. Ferrán (2002) las agrupa en contrastes para una muestra y el resto en no paramétricas. Visauta (2007) engloba todas las pruebas en no paramétricas, mientras que otros autores las clasifican por tipo de muestra. También se aprecia como cada autor recoge un número diferente de pruebas no paramétricas, así como el uso de terminología también diferente para nombrarlas. La intención de este artículo ha sido recoger por primera vez las aportaciones de cada uno de estos autores para ofrecer una clasificación completa de las pruebas no paramétricas (17 en total). En su clasificación se han asumido dimensiones relacionadas con el número de muestras y la relación o independencia entre esas muestras tal y como muestran las tablas 1 y 3. No se han incluido las correlaciones no paramétricas a pesar de que determinados autores las incluyen dentro de pruebas no paramétricas (Pérez Juste *et al.*, 2009), con el objetivo de recopilar solamente los contrastes no paramétricos.

Tabla 1. Resumen de las principales pruebas estadísticas no paramétricas

Variable dependiente	Una muestra (bondad de ajuste)	Muestras relacionadas		Muestras independientes	
		2 muestras	>2 muestras	2 muestras	>2 muestras
Nominal	Binomial Chi-Cuadrado Rachas	McNemar	Cochran	-	-
Ordinal/ Intervalo	Kolmogorov-Smirnov	Signos Wilcoxon	Friedman Kendall	Rachas de Wald-Wolfowitz U de Mann-Whitney Moses Kolmogorov-Smirnov	Mediana Kruskal-Wallis Jonckheere-Terpstra



Se dice que las pruebas no paramétricas son alternativas a las paramétricas, por ello consideramos apropiado establecer la equivalencia entre ambas tal y como se muestra en la tabla 2.

Tabla 2. Pruebas paramétricas y su alternativa no paramétrica

Muestra	Prueba paramétrica	Prueba no paramétrica
<i>Muestras relacionadas</i>		
2 muestras	t-Student	Wilcoxon
> 2 muestras	ANOVA	Friedman
<i>Muestras independientes</i>		
2 muestras	t-Student	U de Mann-Whitney
> 2 muestras	ANOVA	Kruskal-Wallis

En la tabla 3 se recogen las principales características de las 17 pruebas no paramétricas halladas.

PRUEBAS NO PARAMÉTRICAS		
Pruebas	Una muestra	Variables
Prueba de Chi-cuadrado de Pearson	<ul style="list-style-type: none"> Es una prueba de bondad de ajuste, que 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.). 	VD: Nominal
Prueba Binomial	<ul style="list-style-type: none"> Es una prueba de bondad de ajuste, que permite averiguar si una variable dicotómica sigue o no un determinado modelo de probabilidad. Permite contrastar la hipótesis de que la proporción observada de aciertos se ajusta a la proporción teórica de una distribución binomial (lo cual se traduce en la posibilidad de contrastar hipótesis sobre proporciones y sobre cuartiles). 	VD: Nominal
Prueba de Rachas	<ul style="list-style-type: none"> Es una prueba de independencia o de aleatoriedad que permite determinar si el número de rachas (R) observado en una determinada muestra de tamaño n es lo suficientemente grande o lo suficientemente pequeño para poder rechazar la hipótesis de independencia (o aleatoriedad) entre las observaciones. Una racha es una secuencia de observaciones de un mismo atributo o cualidad. Una serie de datos en los que hay muchas o pocas rachas permite concluir que estas no han ocurrido por azar. 	VD: Nominal
Prueba de Kolmogorov-Smirnov (K-S)	<ul style="list-style-type: none"> Es una prueba de bondad de ajuste, que sirve para contrastar la hipótesis nula de que la distribución de una variable se ajusta a una determinada distribución teórica de probabilidad que puede ser con tendencia a la normal, a la de Poisson o exponencial. 	VD: Ordinal/Intervalo
Pruebas	Dos muestras relacionadas	Variables
Prueba de McNemar	<ul style="list-style-type: none"> Sirve para contrastar hipótesis sobre igualdad de proporciones. Se usa cuando hay una situación en la que las medidas de cada sujeto se repiten, por lo que la respuesta de cada uno de ellos se obtiene dos veces: una vez antes y otra después de que ocurra un evento específico. 	VI: Dicotómica VD: Nominal

Vanesa Berlanga y María José Rubio. *Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS*

Prueba de los Signos	<ul style="list-style-type: none"> Permite contrastar la hipótesis de igualdad entre dos medianas poblacionales. Puede ser usada para saber si una variable tiende a ser mayor que otra. También es útil para probar la tendencia que sigue una serie de variables ordinales positivas, o para una valoración rápida de un estudio exploratorio. 	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de Wilcoxon	<ul style="list-style-type: none"> Permite contrastar la hipótesis de igualdad entre dos medianas poblacionales. Paralela a la prueba paramétrica de contraste t para muestras relacionadas. 	VI: Dicotómica VD: Ordinal/Intervalo
Pruebas	K-muestras relacionadas	Variables
Prueba de Friedman	<ul style="list-style-type: none"> Es una extensión de la prueba de Wilcoxon para incluir datos registrados en más de dos periodos de tiempo o grupos de tres o más sujetos pareados, con un sujeto de cada grupo que ha sido asignado aleatoriamente a una de las tres o más condiciones. La prueba examina los rangos de los datos generados en cada periodo de tiempo para determinar si las variables comparten la misma distribución continua de su origen. 	VI: Politémica VD: Ordinal/Intervalo
Prueba de Cochran	<ul style="list-style-type: none"> Esta prueba es idéntica a la prueba de Friedman, pero se aplica cuando todas las respuestas son binarias. La Q de Cochran prueba la hipótesis de que varias variables dicotómicas que están relacionadas entre sí, tienen el mismo promedio. En observaciones múltiples las variables son medidas en el mismo individuo o en individuos pareados. Tiene la ventaja de examinar cambios en las variables categóricas. 	VI: Dicotómica VD: Nominal
Coefficiente de concordancia de W de Kendall	<ul style="list-style-type: none"> Tiene las mismas indicaciones que la prueba de Friedman, aunque su uso en investigación ha sido, principalmente, para conocer la concordancia entre rangos, más que para probar que existe una diferencia entre las medianas. 	VI: Politémica VD: Ordinal /Intervalo
Pruebas	Dos muestras independientes	Variables
Prueba U de Mann-Whitney	<ul style="list-style-type: none"> Es equivalente a la prueba de suma de rangos de Wilcoxon y a la prueba de dos grupos de Kruskal-Wallis. Es la alternativa no paramétrica a la comparación de dos promedios independientes a través de la t de Student. 	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de Kolmogorov-Smirnov	<ul style="list-style-type: none"> 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. 	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de Rachas de Wald-Wolfowitz	<ul style="list-style-type: none"> Contrasta si dos muestras con datos independientes proceden de poblaciones con la misma distribución. 	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de reacciones extremas de Moses	<ul style="list-style-type: none"> Sirve para estudiar si existe diferencia en el grado de dispersión o variabilidad de dos distribuciones. <p>Esta prueba se centra en la distribución del grupo de control y es una medida para saber cuántos valores extremos del grupo experimental influyen en la distribución cuando se combinan con el grupo de control.</p>	VI: Dicotómica VD: Ordinal/Intervalo

Pruebas	K-muestras independientes	Variables
Prueba de la Mediana	<ul style="list-style-type: none"> • Contrasta diferencias entre dos o más grupos en relación con su mediana, bien porque no cumplen las condiciones de normalidad para usar el promedio como medida de tendencia central, bien porque la variable es cuantitativa discreta. • Es similar a la prueba Chi-cuadrado. 	VI: Politémica VD: Ordinal/Intervalo
Prueba de Jonckheere-Terpstra	<ul style="list-style-type: none"> • Es más potente que sus homónimas, la prueba de Kruskal-Wallis y la de la mediana, cuando existe una ordenación <i>a priori</i> (ascendente o descendente) de las K poblaciones de las que se extraen las muestras. 	VI: Politémica VD: Ordinal/Intervalo
Prueba H de Kruskal-Wallis	<ul style="list-style-type: none"> • Es una extensión de la de U de Mann-Whitney y representa una excelente alternativa al ANOVA de un factor completamente aleatorizado. 	VI: Politémica VD: Ordinal/Intervalo

Tabla 3. Resumen de las principales características de las pruebas no paramétricas. Fuente: Elaboración propia.

3. Ejemplificación de las principales pruebas no paramétricas

Para mostrar las principales pruebas no paramétricas en SPSS tomamos como ejemplo un estudio orientado a conocer la actitud del profesorado universitario hacia las TIC en la docencia, a través de una escala tipo Likert. Cada prueba será ejemplificada con un objetivo de investigación adecuado a la prueba, las hipótesis estadísticas correspondientes y la interpretación de los resultados obtenidos en SPSS.

a. Prueba de Chi-cuadrado (una muestra)

Objetivo: Conocer si hay relación entre el sexo y los años de experiencia docente.

H₀: el sexo es independiente de los años de experiencia.

H₁: el sexo y los años de experiencia están relacionados.

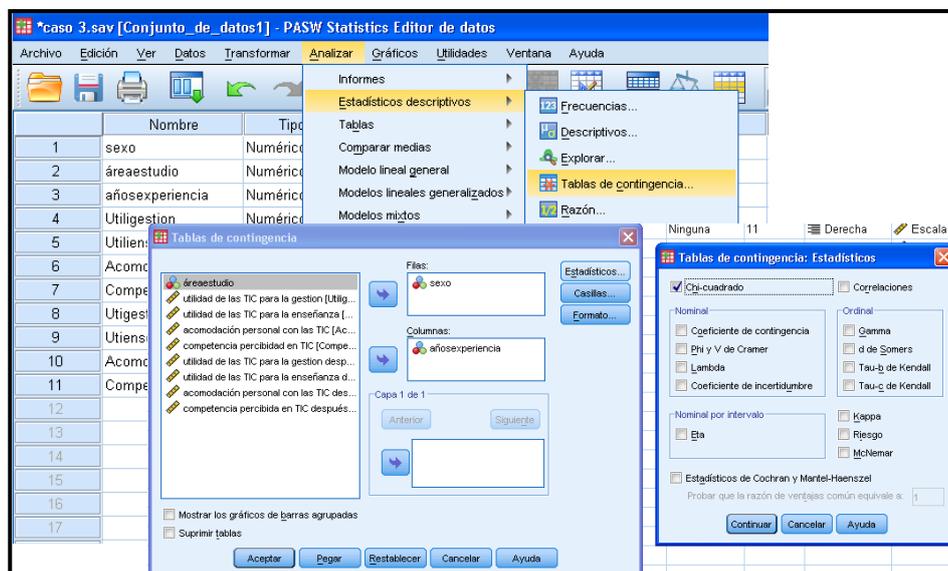


Figura 1. Cuadro de diálogo de la prueba Chi-cuadrado



Resultados

Tabla 4. Output de la prueba Chi-cuadrado

Tabla de contingencia sexo * añosexperiencia

Recuento		añosexperiencia		Total
		menos de 5	5 o más años	
sexo	hombre	18	22	40
	mujer	14	26	40
Total		32	48	80

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,833 ^a	1	,361		
Corrección por continuidad ^b	,469	1	,494		
Razón de verosimilitudes	,835	1	,361		
Estadístico exacto de Fisher				,494	,247
Asociación lineal por lineal	,823	1	,364		
N de casos válidos	80				

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 16,00.

b. Calculado sólo para una tabla de 2x2.

No se rechaza la H_0 , lo que significa que no hay relación entre el sexo y los años de experiencia docente ($\text{sig. } 0,361 > 0,05$).

b. Prueba U de Mann-Whitney (2 muestras independientes)

Objetivo: conocer la influencia de la experiencia docente (<5 años o >5 años) en la utilidad que el profesorado atribuye a las TIC en la enseñanza.

H_0 : la experiencia docente no influye en la utilidad que el profesorado atribuye a las TIC en la enseñanza.

H_1 : la experiencia docente influye en la utilidad que el profesorado atribuye a las TIC en la enseñanza.

Vanesa Berlanga y María José Rubio. *Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS*

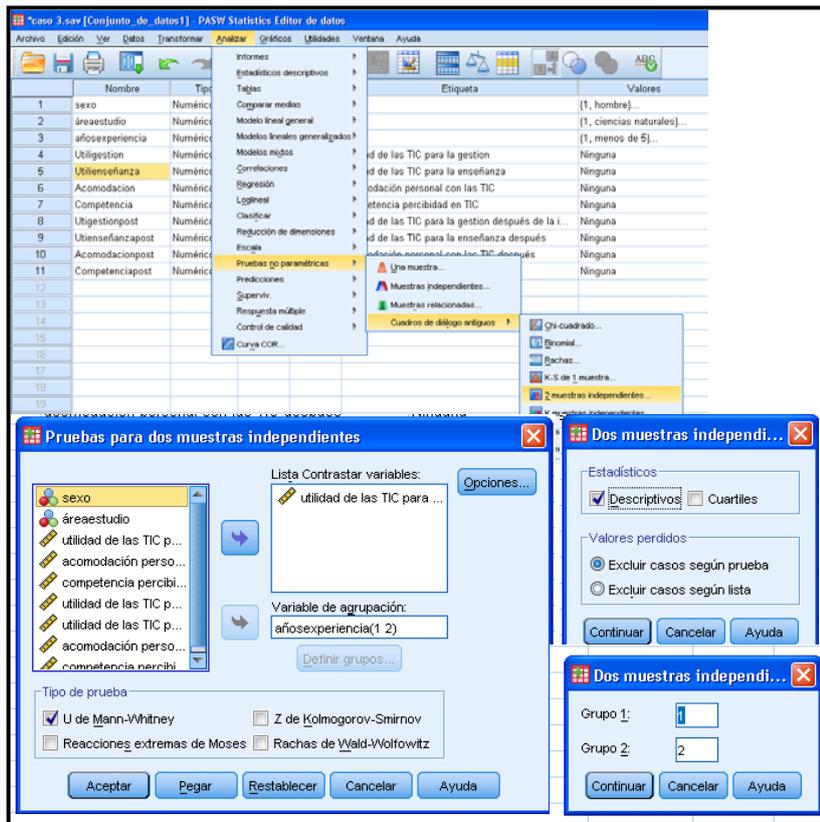


Figura 2. Cuadro de diálogo de la prueba U de Mann-Whitney

Resultados

Tabla 5. Output de la prueba U de Mann-Whitney

Prueba de Mann-Whitney

Rangos				
	añoexperiencia	N	Rango promedio	Suma de rangos
utilidad de las TIC para la enseñanza	menos de 5	32	49,03	1569,00
	5 o más años	48	34,81	1671,00
	Total	80		

Estadísticos de contraste^a

	utilidad de las TIC para la enseñanza
U de Mann-Whitney	495,000
W de Wilcoxon	1671,000
Z	-2,809
Sig. asintót. (bilateral)	,005

a. Variable de agrupación: añoexperiencia

Se acepta la H₁, que significa que la experiencia docente influye en la utilidad que el profesorado atribuye a las TIC en la enseñanza (sig. 0,005<0,05). El profesorado que tiene menos de 5 años de experiencia las considera más útiles.

Vanesa Berlanga y María José Rubio. *Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS*

c. Prueba de Wilcoxon (2 muestras relacionadas)

Objetivo: conocer si los docentes han variado su opinión sobre la utilidad de las TIC en la enseñanza después de la implantación del EEES.

H_0 : no hay diferencias en la opinión de los docentes sobre la utilidad de las TIC en la enseñanza entre antes y después de la implantación del EEES.

H_1 : sí hay diferencias en la opinión de los docentes sobre la utilidad de las TIC en la enseñanza entre antes y después de la implantación del EEES.

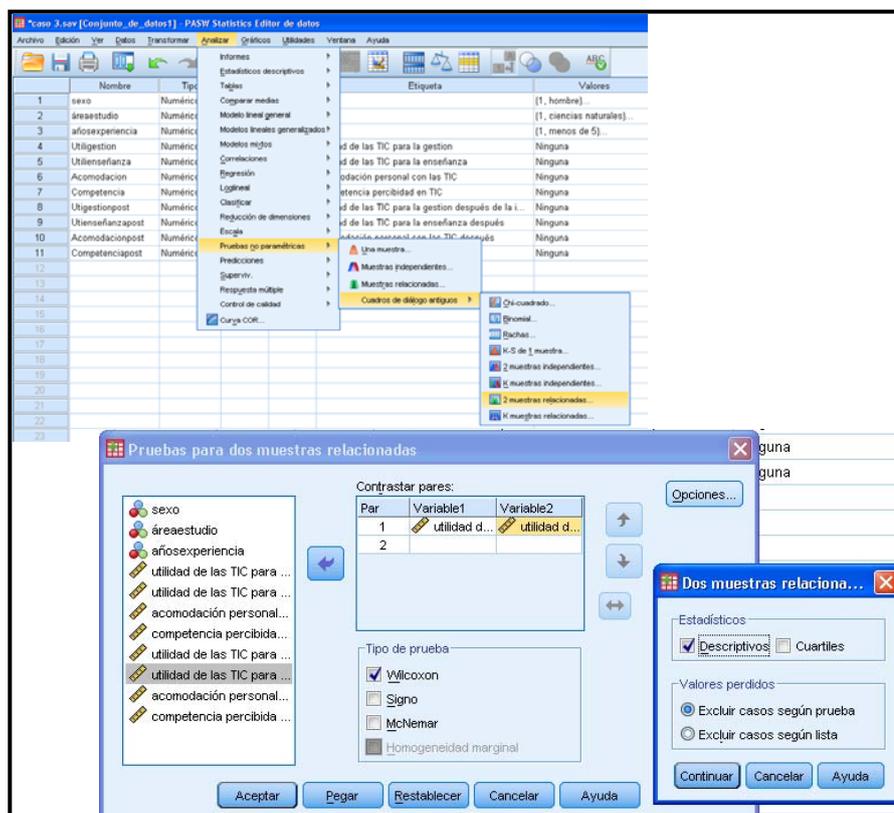


Figura 3. Cuadro de diálogo de la prueba de Wilcoxon



Resultados

Tabla 6. Output de la prueba de Wilcoxon

Prueba de los rangos con signo de Wilcoxon

		Rangos		
		N	Rango promedio	Suma de rangos
utilidad de las TIC para la enseñanza después - utilidad de las TIC para la enseñanza	Rangos negativos	12 ^a	20,50	246,00
	Rangos positivos	33 ^b	23,91	789,00
	Empates	35 ^c		
	Total	80		

a. utilidad de las TIC para la enseñanza después < utilidad de las TIC para la enseñanza

b. utilidad de las TIC para la enseñanza después > utilidad de las TIC para la enseñanza

c. utilidad de las TIC para la enseñanza después = utilidad de las TIC para la enseñanza

Estadísticos de contraste^b

	utilidad de las TIC para la enseñanza después - utilidad de las TIC para la enseñanza
Z	-3,364 ^a
Sig. asintót. (bilateral)	,001

a. Basado en los rangos negativos.

b. Prueba de los rangos con signo de Wilcoxon

Se acepta la H_1 , por lo que concluimos que hay diferencias en la valoración que los docentes hacen de la utilidad de las TIC en la enseñanza entre antes y después de la implantación del EEES (sig.0,001<0,05). El profesorado consideraba más útiles las TIC en la enseñanza antes de la implantación del EEES (23,91).

d. Prueba de Kruskal-Wallis (k muestras independientes)

Objetivo: conocer si el área de estudio (4 áreas) a la que pertenece el profesorado influye en la utilidad que le atribuye a las TIC en la enseñanza.

H_0 : el área de estudio a la que pertenece el profesorado no influye en la utilidad que le atribuye a las TIC en la enseñanza.

H_1 : el área de estudio a la que pertenece el profesorado influye en la utilidad que le atribuye a las TIC en la enseñanza.

Vanesa Berlanga y María José Rubio. *Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS*

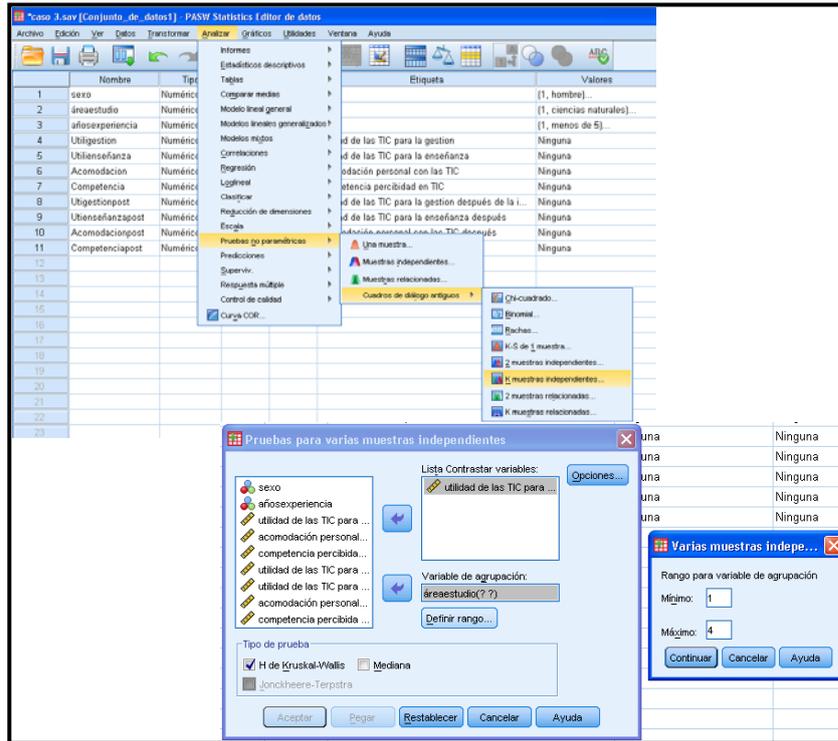


Figura 4. Cuadro de diálogo de la prueba Kruskal-Wallis

Resultados

Tabla 7. Output de la prueba Kruskal-Wallis

Prueba de Kruskal-Wallis

Rangos			
áreaestudio		N	Rango promedio
utilidad de las TIC para la enseñanza	ciencias naturales	20	52,98
	ciencias sociales y educación	20	32,23
	letras	20	20,30
	técnicas	20	56,50
	Total	80	

Estadísticos de contraste^{a, b}

	utilidad de las TIC para la enseñanza
Chi-cuadrado	36,099
gl	3
Sig. asintót.	,000

a. Prueba de Kruskal-Wallis

b. Variable de agrupación: áreaestudio

Vanesa Berlanga y María José Rubio. *Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS*

Se acepta la H_1 y se concluye que el área de estudio a la que pertenece el profesorado influye en la utilidad que le atribuya a las TIC en la enseñanza (sig. $0,000 < 0,05$). El profesorado de las áreas de ciencias naturales (52,98) y de las áreas técnicas (56,50) considera que las TIC son más útiles que el resto de profesorado.

e. Prueba de Friedman (k muestras relacionadas)

Objetivo: conocer si hay diferencias en la valoración que hace el profesorado sobre la utilidad de las TIC en la enseñanza, en la gestión y en el aprendizaje del alumnado.

H_0 : no hay diferencias de valoración entre la utilidad de las TIC en la enseñanza, en la gestión y en el aprendizaje del alumnado.

H_1 : sí hay diferencias de valoración entre la utilidad de las TIC en la enseñanza, en la gestión y en el aprendizaje del alumnado.

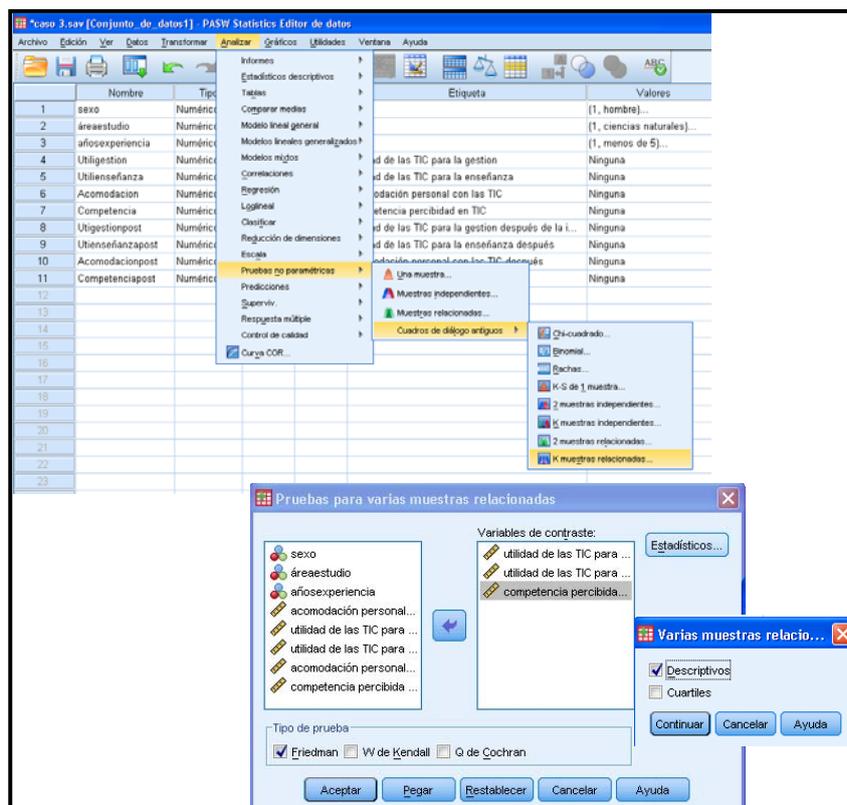


Figura 5. Cuadro de diálogo de la prueba Friedman



Resultados

Tabla 8. Output de la prueba Friedman

Prueba de Friedman

Rangos	
	Rango promedio
utilidad de las TIC para la gestion	1,92
utilidad de las TIC para la enseñanza	2,03
competencia percibida en TIC	2,06

Estadísticos de contraste^a

N	80
Chi-cuadrado	1,750
gl	2
Sig. asintót.	,417

a. Prueba de Friedman

No se rechaza la H_0 y se concluye que el profesorado considera igual de útiles las TIC en la docencia, en el aprendizaje y en la gestión ($\text{sig}.0,417 > 0,05$).



<Referencias bibliográficas>

Ferrán Aranaz, M. (2002) *Curso de SPSS para Windows*. Madrid: McGraw-Hill.

Pérez Juste, R., García Llamas, J.L., Gil Pascual, J.A. y Galán González, A. (2009) *Estadística aplicada a la Educación*. Madrid: UNED - Pearson.

Rubio Hurtado, M. J. y Berlanga Silvente, V. (2012) Cómo aplicar las pruebas paramétricas bivariadas t de Student y ANOVA en SPSS. Caso práctico. En línea *REIRE, Revista d'Innovació i Recerca en Educació*, Vol. 5, núm. 2, 83-100. Accesible en: <http://www.ub.edu/ice/reire.htm>

Visauta Vinacua, B. (2007) *Análisis estadístico con SPSS 14: Estadística básica* (3a ed.). Madrid: McGraw-Hill Interamericana.

Copyright © 2012. Esta obra está sujeta a una licencia de Creative Commons mediante la cual, cualquier explotación de ésta, deberá reconocer a sus autores, citados en la referencia recomendada que aparece al inicio de este documento.

