



Investigación clínica XVI

Diferencias de medianas con la *U* de Mann-Whitney

Rodolfo Rivas-Ruiz,^a Jorge Moreno-Palacios,^a Juan O. Talavera^a

Clinical research XVI. Differences between medians with Mann-Whitney *U* test

If you want to prove that there are differences between two groups with quantitative variables with non-normal distribution, the Mann-Whitney *U* test is used. This test is based on rank differences and it is the opposing of the Student *t* test that use quantitative variables with a normal distribution. If you want to compare three or more non related groups, the Kruskal-Wallis test is applied. When two related samples are compared Wilcoxon test is the best option (a before and after maneuver comparison), when three related samples are compared, Friedman test is used. These test correspond to the parametric opposing paired *t* test and Anova, respectively.

Key words

biomedical research
statistical analysis
reproducibility of results

En la parte XV de esta serie se mostró que la *t* de Student manifiesta las diferencias de medias de dos grupos cuando la variable cuantitativa tiene distribución normal. En este artículo se aborda cómo evaluar una variable ordinal o cuantitativa de libre distribución o, en otras palabras, que no tiene distribución normal. A este grupo de pruebas se le llama “estadística no paramétrica” y la más característica de ellas es la *U* de Mann-Whitney, que contrasta las medianas de dos grupos independientes con libre distribución.

● Ejemplo

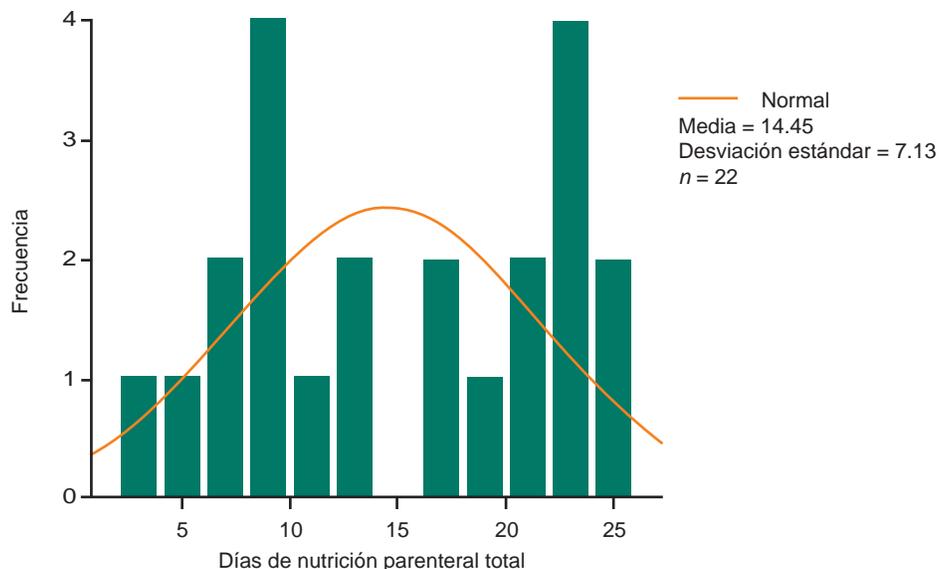
En una unidad de cuidados intensivos neonatales se realiza un estudio para saber si el tiempo de exposición a la nutrición parenteral total se relaciona con el desarrollo de colestasis. Para ello se mide el tiempo de exposición a la nutrición parenteral total en días (variable cuantitativa con libre distribución), la cual constituye la maniobra. Posteriormente se evalúa el desarrollo de la colestasis, que corresponde al desenlace (variable nominal [presente o ausente]). De tal forma, hay una variable cuantitativa de libre distribución y una variable nominal que da lugar a dos grupos.

●●● La hipótesis

El tiempo de uso de la nutrición parenteral total es mayor entre los niños que desarrollan colestasis comparado con los que no la desarrollan.

Para refutar o confirmar esta hipótesis, primero se debe conocer la distribución de la muestra. Para saber si la prueba tiene o no libre distribución se pueden usar las curvas de normalidad (figura 1) y las gráficas Q-Q (figura 2).

Figura 1 Curva de Gauss o curva de normalidad. Una curva de distribución normal debe ser simétrica y descendente a partir de la media (punto mayor), lo cual no ocurre en este ejemplo



Para demostrar que existen diferencias entre grupos independientes con variables cuantitativas que tienen libre distribución, se utiliza la U de Mann-Whitney. Esta prueba tiene su base en la diferencia de rango y es la contraparte de la t de Student que se emplea en las variables cuantitativas con distribución normal. Si se quiere comparar tres o más grupos no relacionados, la prueba que debe aplicarse es la de Kruskal-Wallis. Para contrastar dos muestras relacionadas, la prueba adecuada es la prueba de Wilcoxon (una

comparación de dos momentos: antes y después de una maniobra) y cuando se comparan tres muestras relacionadas debe recurrirse a la de Friedman. Estas pruebas corresponden a sus contrapartes paramétricas t pareada y Anova, respectivamente.

Resumen

Palabras clave

investigación biomédica
análisis estadístico
reproducibilidad de resultados

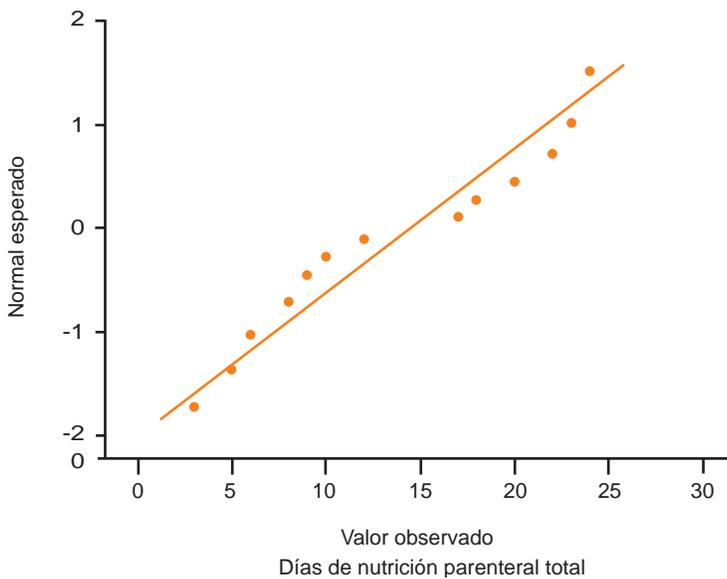


Figura 2 Gráfica Q-Q de normalidad. La línea continua representa la normalidad teórica y los puntos son los valores reales agrupados alrededor de ella. Entre más cercanos estén los puntos de la línea, mayor la posibilidad de que la muestra tenga distribución normal

Tanto la curva de normalidad como la gráfica Q-Q requieren interpretación, la cual puede ser subjetiva. Por ese motivo se calculan otros estadísticos como el sesgo y la curtosis. Para determinar normalidad, ambos valores deben estar en ± 0.5 . Si cualquiera de los valores, ya sea el sesgo o la curtosis, están fuera del rango de ± 0.5 , se asume que la muestra está sesgada y, por lo tanto, tiene una distribución distinta a la normal, es decir, es de libre distribución (cuadro I).

Otra manera de conocer la distribución de la muestra son las pruebas de hipótesis de Kolmogórov-Smirnov y Shapiro-Wilk. Se utiliza la primera si la muestra es mayor de 30 pacientes y la segunda, si está integrada por 30 o menos pacientes. Se asume que la muestra proviene de una población con distribución normal, por lo que la prueba de hipótesis establece lo siguiente:

$H_0 = \text{la muestra} = \text{distribución normal} \rightarrow p > 0.05$

$H_a = \text{la muestra} \neq \text{distribución normal} \rightarrow p \leq 0.05$

Cuadro I Descripción de la muestra, donde aparecen los distintos resultados de los estadísticos y en la parte de abajo el valor del sesgo y la curtosis

	Estadístico	Error estándar
Media	14.455	1.520
Intervalo de confianza para la media de 95 %		
Límite inferior	11.294	
Límite superior	17.616	
Media recortada de 5 %	14.551	
Mediana	14.5	
Varianza	50.831	
Desviación estándar	7.130	
Mínimo	3	
Máximo	24	
Rango	21	
Amplitud intercuartil	14	-0.039
Sesgo	-0.039	0.491
Curtosis	-1.592	0.953

Días de nutrición parenteral total

Cuadro II Resultado de la prueba de normalidad para días de uso de la nutrición parenteral en la UCIN

	Kolmogórov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Días NPT	0.145	22	0.200*	0.904	22	0.036

El grado de libertad (gl) representa $n - 1$, por lo que la muestra se compone de 23 niños

UCIN = unidad de cuidados intensivos neonatales, Sig. = significación verdadera, NPT = nutrición parenteral total

*Límite inferior

Es decir, la muestra es de libre distribución si la p es estadísticamente significativa, como se muestra en el cuadro II (Shapiro-Wilk, $p = 0.036$). Se usa la prueba de Kolmogórov-Smirnov si la población es mayor de 30 o Shapiro-Wilk cuando la muestra es igual o menor de 30.

Una vez que conocemos que la muestra es de libre distribución, se procede a realizar la prueba de contraste de hipótesis, en este caso, la U de Mann-Whitney:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

Donde:

U_1 = muestra con el menor número de pacientes

U_2 = tamaño de muestra mayor

R_1 y R_2 = suma de rangos para cada grupo

La designación n_1 y n_2 es arbitraria si los grupos contienen el mismo número de pacientes. Cada una de las fórmulas da un valor de U distinto (cuadro III).

El primer paso es ordenar de mayor a menor el total de la muestra (ambos grupos); a esto se le llama *rank* o clasificar.

Posteriormente, se calcula la medida de resumen (mediana). Después de ordenar de menor a mayor (5, 7, 9, 12, 18 días), los cinco pacientes del grupo sin colestasis, el que se encuentra a la mitad corresponde a la mediana (valores “rankeados” 1, 3, 4, 6 y 8), en este caso es el “rankeado” con el 4, que corresponde al sujeto con nueve días de exposición a la nutrición parenteral total. De los seis pacientes (2, 5, 7, 9, 10, 11) del grupo con colestasis, la mediana se encuentra entre el 3 y el 4, que corresponden a los valores “rankeados” con 7 y 9 y que corresponden a 15 y 24 días de nutrición parenteral total, los cuales se suman y se dividen entre dos ($29/2$), con lo que se obtiene una mediana de 19.5 días.

Posteriormente se calculan los valores de U_1 y U_2 :

$$U_1 = 22 - \frac{5(5 + 1)}{2} = 15$$

$$U_2 = 44 - \frac{6(6 + 1)}{2} = 21$$

Finalmente, el menor valor de U (en este caso es 15) se coteja en una tabla de valores críticos de alfa (α) de 0.05 para la prueba U de Mann-Whitney, para conocer el valor de p . En este ejemplo, p es de 0.242, es decir, aunque parece haber diferencias en esta pequeña muestra, no alcanzan a ser estadísticamente significativas.

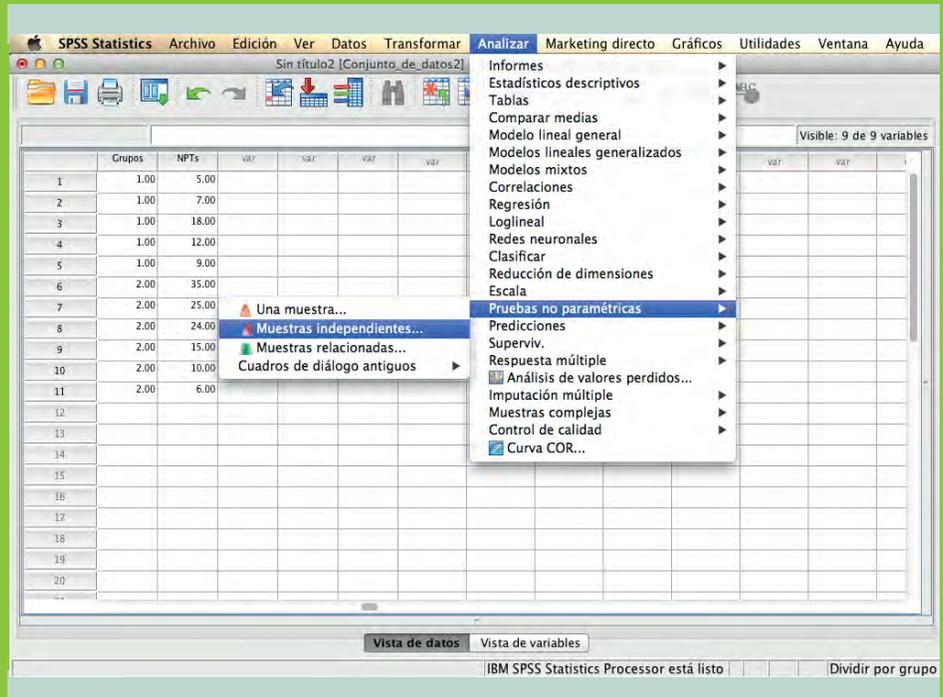
Cuadro III Orden jerárquico de los valores arrojados por los 11 pacientes estudiados en la Unidad de Cuidados Intensivos Neonatales

Sin colestasis $n_1 = 5$	Rank (orden)	Con colestasis $n_2 = 6$	Rank (orden)
5	1	35	11
7	3	25	10
18	8	24	9
12	6	15	7
9	4	10	5
		6	2
Mediana = 9		Mediana = 19.5	
R = 22		R = 44	

Pasos en el programa SPSS

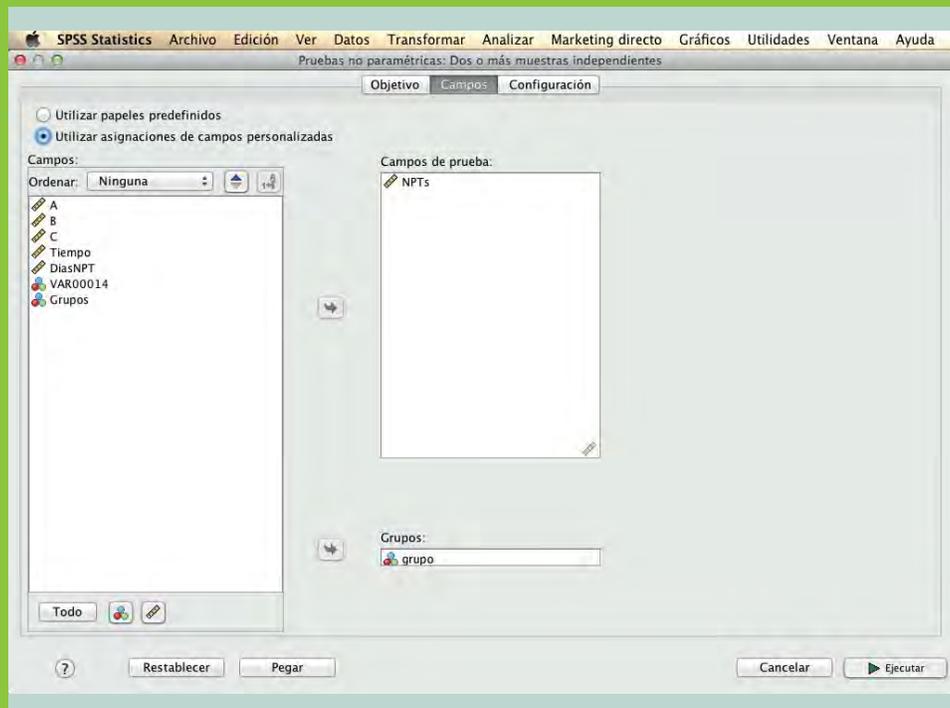
1

En el menú se selecciona la opción “Analizar” y en esta, “Pruebas no paramétricas”. Una vez ahí, se elige “Muestras independientes”.



2

En la siguiente pantalla se solicita ingresar los grupos por contrastar, que en este caso son dos: con colestasis y sin colestasis, y una variable cualitativa, que en este caso es el tiempo de uso de la nutrición parenteral total.



Finalmente, se presiona “Ejecutar” y se obtiene el resultado.

	Hipótesis nula	Test	Sig.	Decisión
1	Las medianas de NPTs son las mismas entre las categorías de grupo.	Prueba de medianas de muestras independientes	.242 ^{1,2}	Retener la hipótesis nula.
2	La distribución de NPTs es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney de muestras independientes	.177 ¹	Retener la hipótesis nula.

Se muestran las significancias asintóticas. El nivel de significancia es .05.

¹Se muestra la significancia exacta para esta prueba.

²Sig. exacta de Fisher

Comentarios

La prueba *U* de Mann-Whitney permite comparar dos medianas, ya sea que provengan de una variable ordinal o de una cuantitativa con libre distribución. Para llegar a esta conclusión será necesario explorar algunos métodos gráficos (imagen de una curva de supervivencia o gráfica Q-Q de normalidad) o algebraicos (sesgo y curtosis o Kolmogórov-Smirnov o Shapiro-Wilk).

Es necesario hacer énfasis en que la *U* de Mann-Whitney es específica para comparar dos muestras

independientes, pero si se trata de tres muestras deberá utilizarse la prueba de Kruskal-Wallis y si se trata de dos muestras relacionadas, la de Wilcoxon; para más de dos muestras relacionadas, la prueba idónea es la de Friedman.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

^aCentro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz
Correo electrónico: rivasrodolfo@gmail.com

Lecturas recomendadas

1. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. Fourth edition. Massachusetts: Blackwell; 2002.
2. Bland M. Introduction to medical statistics. Third edition. New York: Oxford University Press; 2000.
3. Dawson B, Trapp RG. Basic and clinical biostatistics. Fourth edition. New York: Lange Medical Books-McGraw-Hill; 2004.
4. Feinstein AR. Clinical biostatistics. St. Louis: Mosby; 1977.
5. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: Saunders; 1985.
6. Feinstein AR. Multivariable analysis: an introduction. New Haven: Yale University Press; 1996.
7. Le Chap T. Introductory biostatistics. New Jersey: Wiley-Interscience; 2003. Texto libre en [http://www.hstathome.com/tjziyuan/Introductory%20Biostatistics%20Le%20C.T.%20%20\(Wiley,%202003\)\(T\)\(551s\).pdf](http://www.hstathome.com/tjziyuan/Introductory%20Biostatistics%20Le%20C.T.%20%20(Wiley,%202003)(T)(551s).pdf)
8. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Massachusetts: Blackwell; 2005.
9. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. Upper Saddle River, NJ: Pearson, Prentice Hall; 2009.
10. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Norwalk, Conn.: Appleton and Lange; 1993.
11. Rivas-Ruiz R, Pérez-Rodríguez M, Talavera JO. Del juicio clínico al modelo estadístico. Diferencia de medias. Prueba *t* de Student. Rev Med

- Inst Mex Seguro Soc. 2013;51(3):300-3. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=2077:51-3-11-investigacion&catid=785:temas-de-actualidad&Itemid=775
12. Talavera JO, Rivas-Ruiz R. Investigación clínica IV. Pertinencia de la prueba estadística. Rev Med Inst Mex Seguro Soc. 2011;49(4):401-5. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1432:investigacion-clinica-iv-pertinencia-de-la-prueba-estadistica&catid=508:reportes-breves&Itemid=672
 13. Talavera JO, Rivas-Ruiz R. Investigación clínica V. Tamaño de muestra. Rev Med Inst Mex Seguro Soc. 2011;49(5):517-22. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1452:investigacion-clinica-v-tamano-de-muestra&catid=514:temas-de-actualidad&Itemid=676
 14. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Investigación clínica III. Estudios de causalidad. Rev Med Inst Mex Seguro Soc. 2011;49(3): 289-94. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1410:investigacion-clinica-iii-edtudios-de-causalidad&catid=501:temas-de-actualidad&Itemid=664