REVISTA MEXICANA DE ENDOCRINOLOGÍA, METABOLISMO & NUTRICIÓN ARTÍCULO ORIGINAL

Prueba de Wilcoxon-Mann-Whitney: mitos y realidades

*Reinaldo Alberto Sánchez Turcios**

Unidad Médica de Alta Especialidad, Hospital de Cardiología, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

RESUMEN

La prueba de Wilcoxon-Mann-Whitney (WMW) con frecuencia se usa para comparar medias o medianas de dos conjuntos independientes, posiblemente con distribución no normal. Esto no es correcto, y puede conducir a un análisis equívoco. La prueba de WMW establece la diferencia de dispersión de datos de un grupo con respecto a otro. Hacemos énfasis en el uso adecuado de esta prueba. La prueba de WMW es usada como una alternativa para dos muestras independientes de la prueba t de Student. Se hizo un análisis de 100 artículos de publicaciones médicas indexadas y no indexadas nacionales e internacionales, y se observó que la prueba de WMW se usa equívocamente. La prueba de WMW no es exactamente una prueba de medianas, sino que lo importante es la diferencia en la distribución. La prueba de WMW permite encontrar diferentes conclusiones que dependen de la forma y distribución de los datos. Los métodos no paramétricos se basan en el ordenamiento de los datos por su magnitud y en remplazar éstos por la cuantía de los rangos.

Palabras clave: Wilcoxon-Mann-Whitney (WMW). Distribucion no normal. Prueba t-Student para dos muestras independientes.

ABSTRACT

The Wilcoxon-Mann-Whitney (WMW) test is frequently used to compare means and medians of two independent sets with a possible non-normal distribution. This is not correct and can lead to an incorrect analysis. The difference in the data spread of one group compared to another is established by the WMW test. We emphasize the adequate use of this test. The WMW test is used as an alternative of two independent samples of the t-Student test. An analysis was carried out of 100 indexed and non-indexed, medical articles, national, and international, and it was observed that the WMW test was used unequivocally. Thus, the WMW test is not exactly a medians test, but what is important is the spread difference. The WMW test allows us to reach different conclusions that depend on the shape and distribution of data.
The non-parametric methods are based on the ordering of data by magnitude and to replace these data by rank counting.
(REV MEX ENDOCRINOL METAB NUTR. 2015;2:18-21)

Corresponding author: Reinaldo Alberto Sánchez Turcios, rturcios@live.com.mx

Key words: Wilcoxon-Mann-Whitney (WMW). Non-normal distribution. Two independent samples of the t-Student test.

Dirección para correspondencia:
*Reinaldo Alberto Sánchez Turcios
Tepic, 113-610
Col. Roma Sur, C.P.: 06760, México, D.F.
E-mail: rturcios@live.com.mx

Fecha de recepción: 29-09-2014
Fecha de aceptación: 29-01-2015

INTRODUCCIÓN

La prueba de U de Wilcoxon fue diseñada por Frank Wilcoxon entre 1945 y 1950, cuando publicó su artículo original[1]; su enunciado fue mejorado por Henry B. Mann y D. R. Whitney en 1947, con la denominación de prueba para observar una de dos variables aleatorias siendo estocásticamente una más grande que la otra; también proporciona información de un conjunto de datos al compararlo con otro.

La prueba de WMW es conocida como la prueba de suma de rangos y generalmente se usa para comparar las medianas de dos conjuntos independientes. En los últimos años se ha utilizado con mayor frecuencia para comparar dos conjuntos diferentes. Esta prueba se usa en diferentes áreas de la medicina.

La prueba original de WMW se diseñó para probar la hipótesis nula: un elemento de la primera muestra es de menor magnitud con respecto a la segunda y la probabilidad de $p(X < Y) = 0.5$. Sin embargo, la interpretación del valor de p nos permite encontrar evidencia a favor o en contra de la igualdad de las medianas. La falacia en que se basan los que usan esta prueba para comparar la diferencia de medianas es inadecuada. Algunos de los profesionales que usan esta prueba descuidan el uso de parámetros como la media y/o desviación estándar, considerando que prueba de WMW pertenece al grupo de pruebas de la estadística no paramétrica. La popularidad de la prueba ha conducido a errores inconscientes al usarla para comparar la diferencia entre medianas; y la prueba se ha utilizado de forma obstinada a pesar de los nuevos cuestionamientos que han formulado algunas autoridades en estadística. Sólo puede probarse la igualdad de medianas en esta prueba de WMW cuando los dos grupos tienen igualdad en sesgo, curtosis, dispersión. Este procesamiento de datos es referido como modelo de corrimiento puro: se comparan dos histogramas muy semejantes, pero uno de ellos tiene un desplazamiento con respecto al otro; con este modelo la prueba de WMW es robusta para el error tipo I contra la no normalidad y tiene un mayor poder estadístico cuando se compara con otras pruebas, específicamente cuando la distribución de las muestras tiene sesgo (grandes colas de datos a la izquierda o a la derecha de las muestras[2,3]). Varios estudios han demostrado que la prueba de WMW es sensible a las derivaciones del corrimiento puro[4-6]; es necesario mencionar el nivel de significancia. Algunas veces este dato está por debajo y en otros procedimientos está por encima del valor nominal. Actualmente, la prueba de WMW está siendo analizada con criterios muy estrictos. Conover, et al. demostraron que la prueba de WMW es semejante a la t de Student en términos de aceptar o rechazar la hipótesis nula, cuando los datos originales son remplazados por su rango[7]. El problema se estableció ya en el origen cuando Wilcoxon fue incapaz de dar nombre a la prueba y no especificó el tipo de hipótesis que se estaba probando. Fueron H. Mann y D. R. Whitney quienes le dieron el verdadero nombre y además explícitamente la hipótesis aprobada.

MATERIAL Y MÉTODOS

Se examinaron 100 artículos de revistas médicas indexadas, no indexadas, nacionales y extranjeras. En las revistas antes mencionadas hay artículos donde la prueba de WMW se realiza incorrectamente, para comparar medianas en variables continuas con distribución no normal (no gaussiana) y para variables ordinales. Sólo se usa la prueba de WMW para comparar variables continuas por medio de las medianas si uno de los histogramas de las muestras es similar al otro excepto en que se observa un desplazamiento de la parte central hacia la derecha. Si se observaran las gráficas de los dos histogramas y uno de ellos apareciera desplazado a la derecha del otro, y si se sobrepusieran uno al otro, ambos histogramas se observarían muy similares.

Cuando no se da este caso, la prueba de WMW tendrá un valor de p muy significativo, aunque, cuando el valor de las medianas de cada muestra es igual, contradice la hipótesis nula.

Tenemos un ejemplo en la figura 1. Lo que parece una contradicción es que se rechace la hipótesis nula de igualdad de medianas, dado que ambas son iguales a 12 y la p es muy significativa ($p = 0.04463$). Por lo tanto, hay que ser cuidadoso al expresar que se usará la prueba de WMW para encontrar la diferencia en la distribución de los datos de ambas muestras.

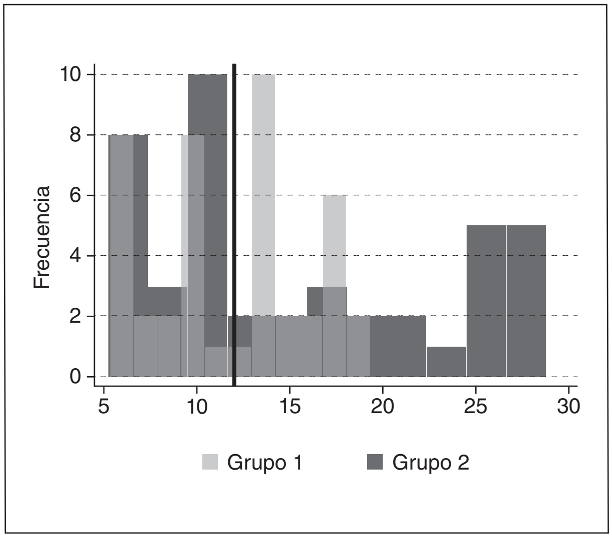

Figura 1. La línea negra es la mediana = 12 para ambas muestras. Con el paquete R se analizaron los datos de los dos conjuntos y se encontraron los resultados de la prueba de WM, donde W = 691.5 y el valor de p = 0.04463[8]; la hipótesis alterna es el desplazamiento de una muestra con respecto a la otra que no es = 0.

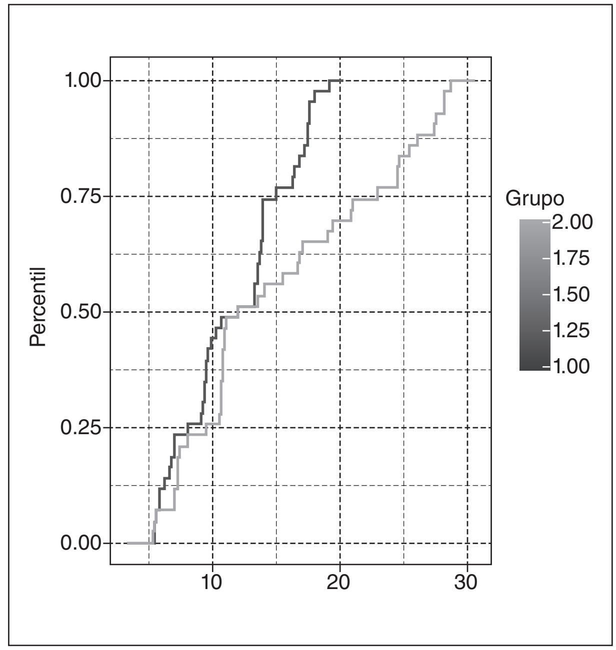

Figura 2.

Naturaleza analítica de la prueba de Wilcoxon-Mann-Whitney

– Se combinan los datos de las dos muestras en una sola y se ordenan de menor a mayor; n1 son los datos de la muestra 1 y n2, los de la muestra 2.

– Se obtienen los rangos que ocupan los datos de la muestra única. Al dato más pequeño se le asigna #1; al que sigue en magnitud, el número 2, donde N = n1 + n2. En el caso de que dos o más datos tengan el mismo valor en el rango, en este proceso se usará el promedio de todos los rangos asignados.

  En esta estructura de datos se debe hacer la anotación del origen de cada dato, es decir, si el dato pertenece a la muestra 1 y/o a la 2. Posteriormente se separan los rangos de la muestra única, determinando a qué muestra pertenecen. De la muestra 1 se obtienen los rangos. Con este procedimiento se obtiene el estadístico U.

– Cálculo del estadístico de la prueba:

$$U = S + 0.5 \times n1 \times (n1 + 1)$$

  Donde S es la suma de los rangos de la muestra n1; en las publicaciones el estadístico U aparece como estadístico T.

– Este valor de U se encuentra en las tablas apropiadas para esta prueba y para obtener así el valor de p.

La prueba de WMW en relación con otras pruebas

Prueba t de Student

Ninguna de las dos muestras tiene distribución normal, el resultado de la t de Student en estos dos grupos es: t = –2.719 y p = 0.008364; por lo tanto, los dos grupos son diferentes en cuanto a la distribución de datos[9] (Fig. 2).

Prueba de Kolmogorov-Smirnov

El estadístico de esta prueba es d = 0.3256 p = 0.02096. Este valor de p es significativo, y nos indica que los datos de cada una de las dos muestras se distribuyen de forma diferente en un grupo con respecto al otro. Esta prueba compara las distribuciones empíricas acumulativas de las dos muestras independientes. Es necesario mencionar que la Academia Americana de Estadística no aconseja usar esta prueba[10].

Prueba de regresión cuantilar

Compara dos grupos mediante el cálculo de los percentiles 25, 50, 75, 90, etc.; en nuestro caso, el percentil 25 da un valor de $p = 0.146$; el 50, $p = 0.999$; el 60, $p = 0.260$; el 65, $p = 0.078$; el 70, $p = 0.003$, y el 80, $p = 0.0001$[11].

DISCUSIÓN

En el 100% de los artículos analizados de revistas médicas con el prestigio de la indexación y no indexación encontramos que la prueba de WMW no se desarrolló con los criterios estadísticos adecuados. Es frecuente observar la confusión en esta prueba comparando medianas, lo cual es un mito; además, muchos artículos establecen el enunciado de que la prueba de WMW pertenece al grupo de pruebas estadísticas paramétricas. En realidad esta prueba determina la probabilidad de que un dato de un grupo sea menor a un dato del otro grupo. La probabilidad constituye la naturaleza de la prueba al analizar el conjunto de datos de los dos grupos, el análisis de la dispersión, el análisis de la fórmula donde se toma en cuenta exclusivamente n1, pero no n2; n2 es útil porque induce el desplazamiento de los números de cada una de las muestras. El análisis de esta prueba y el de otras debe ser una prioridad. Su uso es frecuente en las ciencias biológicas y en medicina, por lo que es importante reexaminar la prueba de WMW. Otra versión es que la prueba de WMW mide sí, la probabilidad de que un dato del primer grupo sea menor que un dato del otro grupo. La t de Student, la prueba de Kolmogorov-Smirnov y la regresión cuantilar determinan la misma diferencia calculada con la prueba de WMW.

CONCLUSIONES

La prueba de WMW pertenece al grupo de pruebas estadísticas no paramétricas. No calcula la diferencia entre las medianas. El cálculo de diferencias de las medianas no es intrínseco en esta prueba. La diferencia de dispersión de datos de un grupo con respecto al otro es la esencia de la prueba de WMW. Un subconjunto de datos establece la diferencia entre los dos grupos.

BIBLIOGRAFÍA

1. Wilcoxon F. Some rapid approximate statistical procedures. Annals of the New York Academy of Sciences. 1950;52:808-14.
2. Bridge PD, Sawilowsky SS. Increasing physicians' awareness of the impact of statistics on research outcomes:comparative power of the t-test and Wilcoxon rank-sum test in small samples applied research. J Clin Epidemiol. 1999;52(3):229-35.
3. Skovlund E, Fenstad GU. Should we always choose a nonparametric test when comparing two apparently nonnormal distributions? J Clin Epidemiol. 2001;54(1):86-92.
4. Hart A. Mann-Whitney test is not just a test of medians: differences in spread can be important. BMJ. 2001;323(7309):391-3.
5. Stonehouse JM, Forrester GJ. Robustness of the t and U tests under combined assumption violations. Journal of Applied Statistics. 1998; 25(1):63-74.
6. Brunner E, Munzel U. The nonparametrics Behrens-Fisher problem: asymptotic theory and a small-sample approximation. Biometrical Journal. 2000;42(1):17-25.
7. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametrics statistics. The American Statistician. 1981;35(3):124-9.
8. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2014. [Internet] Disponible en: http://www.R-project.org/.
9. Wayne WD. Bioestadística: bases para el análisis de las ciencias de la salud. 4.ª ed. México: Limusa; 2002. p. 230-5.
10. Hollander M, Wolfe DA. Nonparametric Statistical Methods. 1.ª ed. Nueva York: John Wiley & Sons; 1973. p. 219-28.
11. Koenker R. Quantile Regression. 1.ª ed. Nueva York: Cambridge University Press; 2005. p. 9-25.