Revista Colombiana de Estadística
Volumen 30 No. 2. pp. 177 a 185. Diciembre 2007

Una exploración de robustez de tres pruebas: dos de permutación y la de Mann-Whitney

Two Permutation Tests and the Mann-Whitney Test: A Robustness Study

Fabián Bautista[a], Emilse Gómez[b]

Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

Resumen

Se exploró y comparó la robustez de las pruebas de Mann-Whitney, de permutación basada en diferencia de medias y de permutación basada en diferencia de medianas mediante simulación para varios grados de violación del supuesto de igualdad de varianzas. Las muestras se generaron de la aproximación a las distribuciones beta, logística y exponencial doble por medio de la familia de distribuciones λ generalizada; se consideraron tamaños de muestra $\{7, 14, 21\}$ y razones entre varianzas $\{0.5, 1.0, 1.5, 2.0\}$. En los casos donde la muestra de menor tamaño proviene de la población con mayor varianza, el nivel de significación simulado toma valores cercanos a tres veces el obtenido cuando se cumple el supuesto. Para la mayoría de los casos estudiados, las pruebas de permutación presentan los mayores niveles de significación simulados.

***Palabras clave*:** simulación, distribución λ generalizada, pruebas no paramétricas, heterocedasticidad.

Abstract

The robustness of the permutation and the Mann-Whitney U tests was explored and compared through simulation for several violation degrees of the variances equality assumption. Samples were obtained from the approximation to beta, logistic and double exponential distributions by means of the Generalized λ Distribution Family; sample sizes $\{7, 14, 21\}$ were considered as well as quotients between variances $\{0.5, 1.0, 1.5, 2.0\}$. When the sample of smaller size comes of the population with greatest variance, the simulated significance level takes values of almost three times the obtained whenever the assumption is fulfilled. For most of the cases, permutation tests present higher simulated significance levels.

***Key words*:** Simulation, Generalized Lambda Distribution, Nonparametric tests, Heteroscedasticity.

[a]Estadístico. E-mail: frbautistar@unal.edu.co
[b]Profesora asistente. E-mail: egomezt@unal.edu.co

1. Introducción

Una de las propiedades deseables para un método estadístico es la robustez; que un procedimiento estadístico sea robusto significa que su desempeño se ve poco o nada afectado por discordancias con el modelo original (presencia de *outliers*, discrepancias en la forma de la densidad en características como el apuntamiento o la simetría) o el no cumplimiento de los supuestos del método; en este caso la estabilidad del desempeño se medirá con el nivel de significancia.

El problema de localización en dos muestras se puede formular de la siguiente manera: se consideran dos muestras aleatorias independientes entre sí $X_1, \dots, X_m$ (primera muestra, muestra de X o primer tratamiento) y $Y_1, \dots, Y_n$ (segunda muestra, muestra de Y o segundo tratamiento); se supone que las muestras provienen de distribuciones absolutamente continuas tales que $P(X_i \leq z) = F(z)$ y $P(Y_j \leq z) = P(X_j \leq z+\delta) = F(z+\delta)$ para todo $i = 1, \dots, m$ y todo $j = 1, \dots, n$. El problema de inferencia es determinar si las dos distribuciones se distinguen en su parámetro de localización, equivalente a probar el sistema de hipótesis: $H_0 : \delta = 0$ frente a $H_0 : \delta \neq 0$.

Las pruebas de permutación, que se exponen en la siguiente sección, y la de Mann-Whitney son pruebas alternativas para el anterior problema, que pueden tener idénticos supuestos: las funciones de distribución son absolutamente continuas, tienen la misma forma y solo difieren en su parámetro de localización. En particular, si se usa la familia de distribuciones λ generalizada, esto implica que tienen las mismas varianzas y parámetros de asimetría y kurtosis. Este estudio de robustez examinó el desempeño de las pruebas cuando no se cumple el supuesto de igualdad de varianzas, mediante la comparación de sus niveles de significación calculados vía simulación; es de notar que este tema ha sido abordado por autores como Sortres-Ramos & Castillo-Márquez (2000) y Stonehouse & Forrester (1998) para la prueba basada en rangos, pero la documentación disponible acerca de esta propiedad para pruebas de permutaciones es escasa, salvo el artículo recientemente publicado por Neubert & Brunner (2007) que utiliza una estadística para dos muestras con varianzas distintas.

Para comprender los alcances de este estudio, a continuación se presentan algunos conceptos básicos de las pruebas de permutación para la diferencia en localización de dos muestras independientes, luego los resultados con su metodología de obtención y por último las conclusiones extraídas.

2. Pruebas de permutación para dos muestras

La ausencia de muestras aleatorias en algunos campos del conocimiento, por ejemplo en investigación biomédica (Ludbrook & Dudley 1998), obliga a recurrir a métodos de aleatorización en la asignación de los sujetos experimentales a los dos tratamientos con objeto de hacer inferencias estadísticas acerca de los cambios en alguno de los parámetros de la variable medida. En particular si el investigador está interesado en lograr un mejor desempeño reflejado en la tendencia central de la variable, la inferencia se hará a través de una comparación entre las medias de los dos tratamientos; la hipótesis alternativa se define según si la característica medida hace referencia a un incremento (por ejemplo, en caso de pretender aumentar producción de glóbulos rojos) o a una disminución (por ejemplo, cuando se busca bajar el nivel de glucosa).

El argumento es el siguiente: si los dos tratamientos no generan efectos diferentes en la variable respuesta, entonces todos los $\binom{m+n}{n}$ conjuntos de datos obtenidos al asignar aleatoriamente los sujetos a los tratamientos tendrían la misma probabilidad de ser observados. La mayoría de estos conjuntos de datos tendrán observaciones grandes y pequeñas asignadas a cada tratamiento. Estos conjuntos de datos serían los que se esperarían observar si los dos tratamientos fueran iguales; en tal caso, la media de los dos tratamientos sería parecida. Por otro lado, si el primer tratamiento es más efectivo, entonces las observaciones más grandes tenderían a ocurrir en el primer tratamiento; en consecuencia, la media del primer tratamiento sería mayor que la media del segundo tratamiento y la diferencia sería más grande de la que se esperaría ocurriera por azar.

Este argumento para probar la diferencia entre medias también aplica para la diferencia entre medias recortadas u otras estadísticas (Higgins 2004); además el fundamento de intercambiabilidad se puede adaptar para otros problemas de inferencia como la diferencia entre medianas o diferencia entre parámetros de escala. En general, el investigador tiene libertad para escoger la estadística que considere mejor para describir la diferencia entre dos grupos en términos del parámetro de interés y luego usar una prueba de permutación para determinar si la diferencia es significativa o no.

Las pruebas de permutación para dos muestras independientes fundamentan su rechazo de la hipótesis de igualdad de distribuciones en favor de la alternativa específica, si el nivel de significancia predeterminado es mayor que el correspondiente valor p. En el problema de interés, diferencia en localización, el procedimiento para obtener un valor p aproximado para la prueba de permutación basada en la diferencia de medias (extensible a medianas) se describe a continuación:

1. Asignar aleatoriamente las unidades experimentales (sujetos u objetos) a uno de los dos tratamientos: m al primero y n al segundo. Obtener la información de las unidades y calcular la diferencia entre las dos medias, $D_{\text{obs}} = \overline{X}_m - \overline{Y}_n$.
2. Crear un vector con las $m + n$ observaciones.
3. Sortear aleatoriamente los elementos de este vector, asignando los primeros m al primer tratamiento y los restantes n al segundo.
4. Calcular la diferencia entre medias del conjunto de datos sorteados, $D_1 = (\overline{X}_m - \overline{Y}_n)_1$.
5. Repetir los dos pasos anteriores R veces, garantizando una muestra aleatoria con remplazo de todas las permutaciones de los elementos. El resultado hasta este paso será el conjunto de diferencias $D_1, \dots, D_R$.
6. Si el efecto presumido del primer tratamiento es producir observaciones más grandes en promedio que las del segundo (prueba unilateral de cola superior),

calcular el valor p aproximado como la fracción de las diferencias entre medias sorteadas que sean mayores o iguales a la diferencia observada, D_{obs}; es decir:

$$p_{\text{cola superior}} = \frac{\#\{D_i \geq D_{\text{obs}}\}}{R}, \qquad i = 1, \dots, R$$

En forma análoga, si la prueba unilateral es de cola inferior, calcular la fracción de los D_i menores o iguales a D_{obs}.

Si la prueba es bilateral, realizar un procedimiento análogo con el valor absoluto de las diferencias entre las medias; es decir:

$$p_{\text{dos colas}} = \frac{\#\{|D_i| \geq |D_{\text{obs}}|\}}{R}, \qquad i = 1, \dots, R$$

3. Resultados

Los niveles de significación exactos requeridos para estudiar la robustez de las pruebas de permutación y U de Mann-Whitney no se pueden obtener teóricamente de una manera sencilla, por tanto, se calcularon por medio de una simulación cuyas características se comentan a continuación:

Distribuciones. Se seleccionaron tres casos de la familia de distribuciones λ generalizada para generar los conjuntos de datos. Emplear esta familia facilitó generar distribuciones con los mismos parámetros de forma (asimetría y kurtosis) y localización (media), pero con parámetros de escala distintos (varianzas distintas); además del cubrimiento de formas distintas de modo que para cada par de muestras las distribuciones fueran casi iguales.

Se utilizaron las aproximaciones a las distribuciones beta con $\beta_3 = \beta_4 = 1$ (distribución uno), logística con $\mu = 0$ y $\sigma = 1$ (distribución dos) y exponencial doble con $\lambda = 1$ (distribución tres), según las definiciones en Karian & Dudewicz (2000) (véase la tabla 1).

Pruebas. Como las distribuciones seleccionadas son simétricas, la media es igual a la mediana en valor; de esta manera se pueden utilizar las pruebas U y de permutación usando la mediana como pruebas alternativas a la de permutación usando la media.

Tabla 1: Parámetros de la familia de distribuciones λ generalizada para aproximar las distribuciones.

Distribución	λ_1	λ_2	λ_3	λ_4
Beta	0.5	1.9693000	0.449500	0.449500
Logística	0	-0.0003637	-0.000363	-0.000363
Exponencial doble	7.5505×10^{-17}	-0.1192000	-0.080200	-0.080200

Tamaños de muestras. Los pares de tamaños de muestras usados fueron: $(n_1, n_2) \in \{(7,7), (14,14), (21,21), (7,14), (7,21), (14,21)\}$, los cuales permiten trabajar con las distribuciones exactas del estadístico U de Mann-Whitney y existen diferentes proporciones entre ellos.

Relación entre varianzas. Para todas las distribuciones y tamaños de muestras se seleccionaron parámetros asociados a la escala de la distribución λ generalizada de tal forma que la relación entre varianzas fuera $\sigma_1^2/\sigma_2^2 = 0.5$, $\sigma_1^2/\sigma_2^2 = 1.5$ y $\sigma_1^2/\sigma_2^2 = 2$. Además se utilizó la relación $\sigma_1^2/\sigma_2^2 = 1$; en otros términos, se calcularon los niveles de significación cuando se cumple el supuesto de igualdad de varianzas, para poder comparar los niveles de significación ante la violación del supuesto de igualdad de varianzas, ya que para la prueba U se conocen los niveles de significación nominales, pero para las pruebas de permutación no.

Para cada elección de distribución, par de tamaños de muestra y relación entre varianzas se simuló el nivel de significación de cada prueba mediante un programa escrito en R (R Development Core Team 2007) versión 2.5.1 (ver tabla 2). La prueba U se hizo utilizando la función `wilcox.test`, mientras que las pruebas de permutación se programaron usando una muestra de 1000 permutaciones para hallar el valor p aproximado. Cada prueba se hizo a dos colas con nivel de significación teórico de 5 %.

El algoritmo de simulación se presenta a continuación:

1. Generar conjuntos de números pseudo-aleatorios de poblaciones que tengan los mismos parámetros de forma y localización según las distribuciones, tamaños de muestras y relaciones entre varianzas anteriores.

2. Realizar las pruebas de Mann-Whitney (U), de permutación usando la media (M) y la mediana (N) a los conjuntos de datos simulados y guardar la decisión del rechazo o no rechazo de la hipótesis nula.

3. Realizar los pasos anteriores 1000 veces.

4. Contar el número de rechazos de la hipótesis nula; en otras palabras, la cantidad de errores de tipo I, ya que las muestras vienen de poblaciones con el mismo parámetro de localización.

La prueba más robusta es aquella cuyo nivel de significación simulado sea más cercano al obtenido cuando se cumple el supuesto de igualdad de varianzas, aproximadamente 5 %; para analizar esto de una manera sencilla, se calcularon los alejamientos entre niveles de significación simulados como la diferencia entre los valores cuando la relación entre varianzas toma el valor uno y cuando toma otro valor (ver tabla 3).

4. Discusión y conclusiones

En todos los tamaños de muestras, aproximaciones a las distribuciones y pruebas, los alejamientos más grandes se presentan la mayoría de las veces cuando $\sigma_1^2/\sigma_2^2 = 0.5$ o $\sigma_1^2/\sigma_2^2 = 2$; estas relaciones entre varianzas son análogas en el sentido que una población tiene el doble de varianza que la otra. Por ejemplo, cuando las muestras provienen de la *distribución tres* (exponencial doble) en los tamaños de muestras iguales a siete, los niveles de significación simulados de la prueba U

Tabla 2: Niveles de significación simulados.

	Distribución uno Beta			Distribución dos Logística			Distribución tres Exponencial doble		
Parámetros	U	M	N	U	M	N	U	M	N
$n = m = 7$									
$\sigma_1^2/\sigma_2^2 = 1$	3.6	4.5	3.6	4.1	5.1	4.2	4.6	5.7	3.5
$\sigma_1^2/\sigma_2^2 = 0.5$	4.2	5.7	4.8	3.6	4.4	3.8	4.0	4.6	3.9
$\sigma_1^2/\sigma_2^2 = 1.5$	3.9	5.2	4.0	4.1	4.9	4.0	4.3	5.5	3.8
$\sigma_1^2/\sigma_2^2 = 2$	4.0	4.5	3.8	3.8	5.0	2.9	3.4	5.0	3.9
$n = m = 14$									
$\sigma_1^2/\sigma_2^2 = 1$	5.2	5.6	5.5	4.4	4.4	5.2	4.8	5.6	4.1
$\sigma_1^2/\sigma_2^2 = 0.5$	6.1	5.8	7.6	3.1	3.4	5.8	4.8	5.4	5.8
$\sigma_1^2/\sigma_2^2 = 1.5$	5.0	4.2	5.5	4.7	5.2	5.6	4.8	4.9	5.2
$\sigma_1^2/\sigma_2^2 = 2$	6.2	5.6	6.5	5.3	5.5	6.0	4.7	5.7	4.9
$n = m = 21$									
$\sigma_1^2/\sigma_2^2 = 1$	3.6	3.7	4.0	4.6	5.1	4.9	4.7	5.6	5.1
$\sigma_1^2/\sigma_2^2 = 0.5$	4.6	5.3	5.9	5.1	4.9	6.6	5.3	5.6	5.2
$\sigma_1^2/\sigma_2^2 = 1.5$	5.5	5.7	6.3	5.1	5.7	5.8	5.1	5.3	5.6
$\sigma_1^2/\sigma_2^2 = 2$	4.9	4.6	4.7	6.1	7.0	6.7	5.6	5.5	6.6
$n = 7,\ m = 14$									
$\sigma_1^2/\sigma_2^2 = 1$	4.8	4.6	5.5	4.7	5.0	5.2	6.6	6.8	6.0
$\sigma_1^2/\sigma_2^2 = 0.5$	3.0	2.3	2.2	2.3	2.1	2.5	3.1	2.0	2.9
$\sigma_1^2/\sigma_2^2 = 1.5$	6.3	5.8	7.5	5.5	6.5	6.4	4.5	5.8	5.9
$\sigma_1^2/\sigma_2^2 = 2$	6.0	7.4	9.8	5.9	7.5	9.0	6.5	9.5	9.1
$n = 7,\ m = 21$									
$\sigma_1^2/\sigma_2^2 = 1$	4.3	4.0	3.3	5.6	5.6	5.0	4.9	5.3	5.3
$\sigma_1^2/\sigma_2^2 = 0.5$	2.3	1.7	1.9	3.1	2.9	1.9	3.4	3.1	1.9
$\sigma_1^2/\sigma_2^2 = 1.5$	6.5	7.7	7.4	7.2	8.7	7.2	6.4	7.4	7.9
$\sigma_1^2/\sigma_2^2 = 2$	6.4	9.6	11.6	8.9	11.9	12.0	7.2	10.9	10.2
$n = 14,\ m = 21$									
$\sigma_1^2/\sigma_2^2 = 1$	4.2	4.5	5.3	5.5	5.4	5.8	5.7	5.7	5.1
$\sigma_1^2/\sigma_2^2 = 0.5$	3.6	3.8	4.1	4.0	3.7	4.0	4.2	3.7	3.9
$\sigma_1^2/\sigma_2^2 = 1.5$	5.9	6.1	6.4	6.1	7.7	6.8	5.7	6.8	6.2
$\sigma_1^2/\sigma_2^2 = 2$	8.4	8.3	9.4	6.8	7.3	7.9	6.8	7.3	8.7

están por debajo del obtenido cuando se cumple el supuesto de igualdad de varianzas, es decir, la prueba se vuelve conservativa; los alejamientos son de 0.6 %, 0.3 % y 1.2 % para las relaciones entre varianzas en aumento.

En todos los tamaños de muestras y aproximaciones a las distribuciones las pruebas de permutación presentan los mayores alejamientos para la mayoría de las relaciones entre varianzas. En representación de esto, cuando las muestras provienen de la *distribución dos* (logística), $n = 7$, $m = 21$ y $\sigma_1^2/\sigma_2^2 = 2$, los alejamientos son 3.3 % para la prueba U, 6.3 % para la de permutación usando la media y 7 % para la que usa la mediana; correspondiendo a niveles de significación simulados de 8.9 %, 11.9 % y 12 %.

Cuando los tamaños de muestras son distintos, los alejamientos son mayores debido a que se alcanzan niveles de significación simulados de hasta 12 %; en tanto los niveles de significación calculados en los tamaños de muestras iguales no están muy lejos del nivel de significación deseado, es decir, de 5 %. Por consiguiente, se

TABLA 3: Alejamientos.

	Distribución uno Beta			Distribución dos Logística			Distribución tres Exponencial doble		
Parámetros	U	M	N	U	M	N	U	M	N
$n = m = 7$									
$\sigma_1^2/\sigma_2^2 = 0.5$	−0.6	−1.2	−1.2	+0.5	+0.7	+0.4	+0.6	+1.1	−0.4
$\sigma_1^2/\sigma_2^2 = 1.5$	−0.3	−0.7	−0.4	0.0	+0.2	+0.2	+0.3	+0.2	−0.3
$\sigma_1^2/\sigma_2^2 = 2$	−0.4	0.0	−0.2	+0.3	+0.1	+1.3	+1.2	+0.7	−0.4
$n = m = 14$									
$\sigma_1^2/\sigma_2^2 = 0.5$	−0.9	−0.2	−2.1	+1.3	+1.0	−0.6	0.0	+0.2	−1.7
$\sigma_1^2/\sigma_2^2 = 1.5$	+0.2	+1.4	0.0	−0.3	−0.8	−0.4	0.0	+0.7	−1.1
$\sigma_1^2/\sigma_2^2 = 2$	−1.0	0.0	−1.0	−0.9	−1.1	−0.8	+0.1	−0.1	−0.8
$n = m = 21$									
$\sigma_1^2/\sigma_2^2 = 0.5$	−1.0	−1.6	−1.9	−0.5	+0.2	−1.7	−0.6	0.0	−0.1
$\sigma_1^2/\sigma_2^2 = 1.5$	−1.9	−2.0	−2.3	−0.5	−0.6	−0.9	−0.4	+0.3	−0.5
$\sigma_1^2/\sigma_2^2 = 2$	−1.3	−0.9	−0.7	−1.5	−1.9	−1.8	−0.9	+0.1	−1.5
$n = 7,\ m = 14$									
$\sigma_1^2/\sigma_2^2 = 0.5$	+1.8	+2.3	+3.3	+2.4	+2.9	+2.7	+3.5	+4.8	+3.1
$\sigma_1^2/\sigma_2^2 = 1.5$	−1.5	−1.2	−2.0	−0.8	−1.5	−1.2	+2.1	+1.0	+0.1
$\sigma_1^2/\sigma_2^2 = 2$	−1.2	−2.8	−4.3	−1.2	−2.5	−3.8	+0.1	−2.7	−3.1
$n = 7,\ m = 21$									
$\sigma_1^2/\sigma_2^2 = 0.5$	+2.0	+2.3	+1.4	+2.5	+2.7	+3.1	+1.5	+2.2	+3.4
$\sigma_1^2/\sigma_2^2 = 1.5$	−2.2	−3.7	−4.1	−1.6	−3.1	−2.2	−1.5	−2.1	−2.6
$\sigma_1^2/\sigma_2^2 = 2$	−2.1	−5.6	−8.3	−3.3	−6.3	−7.0	−2.3	−5.6	−4.9
$n = 14,\ m = 21$									
$\sigma_1^2/\sigma_2^2 = 0.5$	+0.6	+0.7	+1.2	+1.5	+1.7	+1.8	+1.5	+2.0	+1.2
$\sigma_1^2/\sigma_2^2 = 1.5$	−1.7	−1.6	−1.1	−0.6	−2.3	−1.0	0.0	−1.1	−1.1
$\sigma_1^2/\sigma_2^2 = 2$	−4.2	−3.8	−4.1	−1.3	−1.9	−2.1	−1.1	−1.6	−3.6

puede sugerir la aplicación de estas pruebas, especialmente la prueba U, ante la violación del supuesto de igualdad de varianzas cuando los tamaños de muestras son iguales.

Por ejemplo, cuando las muestras provienen de la *distribución dos* (logística) y $n = m = 14$, el nivel de significación simulado para la prueba de permutación usando la mediana, si se cumple el supuesto de igualdad de varianzas, es 5.2 %, mientras que el calculado, si $\sigma_1^2/\sigma_2^2 = 2$ es 6 %, es decir, un alejamiento de 0.8 %; cuando $n = 7$ y $m = 21$, el nivel de significación simulado si se cumple el supuesto es 5 %, mientras que el obtenido si $\sigma_1^2/\sigma_2^2 = 2$ es 12 %, es decir, un alejamiento de 7 %.

En los tamaños de muestras distintos las pruebas se vuelven conservativas cuando $\sigma_1^2/\sigma_2^2 = 0.5$, es decir, cuando una muestra tiene el doble de varianza y el doble o triple de tamaño de muestra que la otra. En adición, esta característica se observa para la aproximación a la exponencial doble en los casos $n = 7$, $m = 14$ y $\sigma_1^2/\sigma_2^2 = 1.5$ y en la relación $\sigma_1^2/\sigma_2^2 = 2$ solo en la prueba U.

Los niveles de significación simulados más altos se presentan en los tamaños de muestras distintos cuando $\sigma_1^2/\sigma_2^2 = 2$, es decir, cuando la muestra con la mayor varianza es la de menor tamaño, sobre todo para las pruebas de permutación. En consecuencia, sería razonable realizar alguna transformación a los datos para homogeneizar varianzas y luego aplicar las pruebas.

Las dos características anteriores presentadas en los tamaños de muestras distintos, junto con que los mayores alejamientos se presentan en las relaciones $\sigma_1^2/\sigma_2^2 = 0.5$ y $\sigma_1^2/\sigma_2^2 = 2$, hacen que en estos tamaños de muestras los niveles de significación simulados tiendan a crecer a medida que la relación entre varianzas aumenta. Para ilustrar el comentario anterior, cuando las muestras provienen de la *distribución tres* (exponencial doble) con $n = 7$ y $m = 21$, los niveles de significación de la prueba de permutación usando la media son 3.1 %, 5.3 %, 7.4 % y 10.9 % para las cuatro relaciones entre varianzas en orden ascendente; correspondiendo a alejamientos de 2.2 % por debajo del nivel de significación obtenido si se cumple el supuesto de igualdad de varianzas; 2.1 % y 5.6 % por encima del simulado si se cumple el supuesto.

En términos de las tres distribuciones no se observan mayores cambios entre una y otra, ni en términos de la estadística usada en las pruebas de permutación (media o mediana); esto puede estar relacionado con la característica común de simetría en las distribuciones. Esta similitud en los comportamientos de las pruebas de permutación sugiere que el cambio en el estadístico empleado no es un elemento decisivo en la selección de la prueba no paramétrica siempre que éste se relacione con el parámetro de interés.

Las tres pruebas comparadas pierden idoneidad cuando se incumple el supuesto, resaltando que las pruebas de permutación presentan los mayores alejamientos, en especial cuando los tamaños de muestras son distintos. En consecuencia, no se recomienda el uso de estas pruebas si la varianza difiere entre las muestras; en tal caso, se propone una prueba de permutaciones con una estadística que considere la heterocedasticidad o la transformación previa de los datos con el fin de homogeneizar varianzas.

Finalmente, la continuidad de este tipo de estudios o el desarrollo de fundamentos teóricos respecto a la bondad de las pruebas de permutación es un campo de investigación que genera grandes aportes con relación a la disminución de errores y el aumento en niveles de confiabilidad ocasionados por el uso inadecuado de técnicas estadísticas. Cabe recordar que los resultados obtenidos vía simulación no se pueden extrapolar a contextos con números de observaciones distintas o cuando las distribuciones generadoras de los datos no sean continuas ni con formas aproximadas a los tres casos considerados.

Recibido: febrero de 2007
Aceptado: agosto de 2007

Referencias

Higgins, J. J. (2004), *An Introduction to Modern Nonparametric Statistics*, Brooks/Cole, Pacific Grove, USA.

Karian, Z. & Dudewicz, E. (2000), *Fitting Statistical Distributions to Data: The Generalized Lambda Distribution and the Generalized Bootstrap Methods*, Boca Raton, USA.

Ludbrook, J. & Dudley, H. (1998), 'Why Permutation Tests are Superior to t and F Tests in Biomedical Research', *The American Statistician* **52**(2), 127–132.

Neubert, K. & Brunner, E. (2007), 'A Studentized Permutation Test for the Non-parametric Behrens-Fisher Problem', *Computational Statistics and Data Analysis* **51**, 5192–5204.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org

Sortres-Ramos, D. & Castillo-Márquez, L. E. (2000), 'Estimación del nivel de significancia real de la prueba de Mann-Whitney ante violaciones a los supuestos estándar usando simulación Montecarlo', *Agrociencia* **34**(1), 69–74.

Stonehouse, J. M. & Forrester, G. J. (1998), 'Robustness of the t and U Tests Under Combined Assumption Violations', *Journal of Applied Statistics* **1**, 63–74.