

Mann–Whitney U-test. Alternativa no paramétrica al t-test de medias independientes

Joaquín Amat Rodrigo

Enero, 2016

Índice

Introducción.....	2
Ventajas y desventajas de <i>U de Mann- Withney</i> frente a <i>t-test</i>	2
Condiciones necesarias del U-test.....	5
Ejemplo.....	5
1.Hipotesis:	6
2.Análisis de homocedasticidad.....	6
3.Ordenación de las observaciones	6
4.Cálculo del estadístico U	7
5.Cálculo de p-value	8
6.Cálculo del tamaño del efecto (side effect)	8
Solución directa con R	9
Resultado.....	10
Test de Kolmogorov-Smirnov	10

Introducción

Ventajas y desventajas de *U de Mann- Withney* frente a *t-test*

El test no paramétrico *U-test de Mann-Withney*, también conocido como *Mann-Whitney-Wilcoxon (MWW)*, *Wilcoxon rank-sum test*, o *Wilcoxon-Mann-Whitney test* permite comparar poblaciones cuando sus distribuciones (normalmente interpretadas a partir de las muestras) no satisfacen las condiciones necesarias para otros test paramétricos. Son una alternativa a los *t-test para muestras independientes* cuando las muestras no siguen una distribución normal (muestran asimetría o colas) o porque tienen un tamaño demasiado reducido para poder determinar si realmente las poblaciones de las que proceden se distribuyen de forma normal.

A la hora de elegir entre *t-test* o *U-test*, es importante tener en cuenta que el problema de las muestras pequeñas no se soluciona con ninguno de los dos. Si el tamaño de las muestras es pequeño también lo es la calidad de la inferencia que se puede hacer con ellas. Ahora bien, existen dos situaciones en las que *a priori* se puede recomendar utilizar un *U-test* antes que un *t-test*:

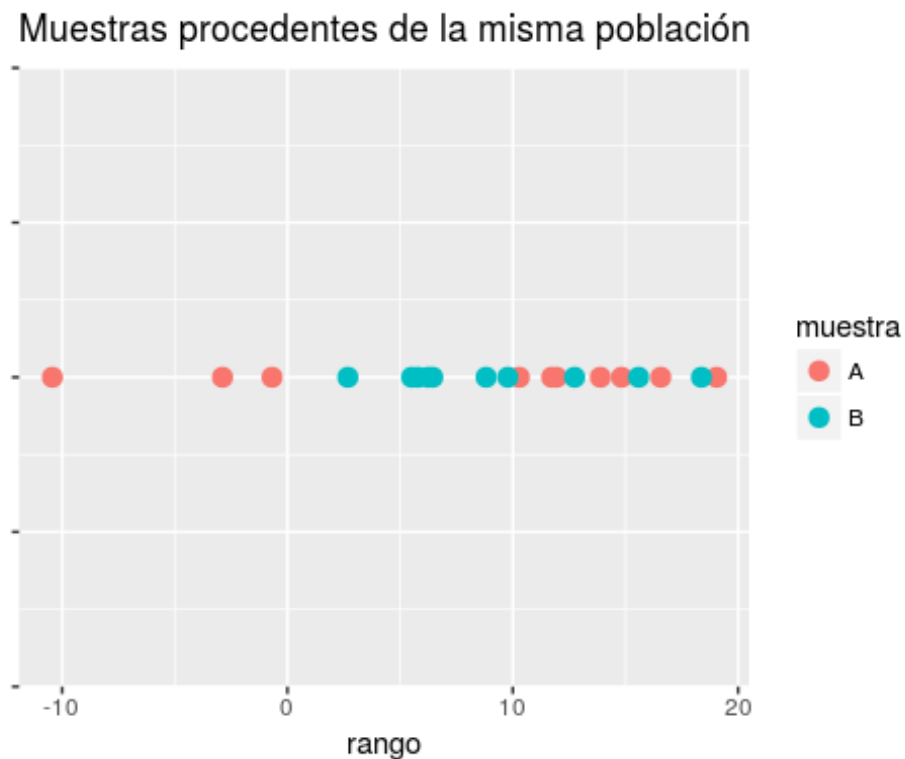
- Si el tamaño de las muestras es suficientemente grande para determinar (por métodos gráficos o test analíticos) que la distribución de las poblaciones a comparar no es de tipo normal, en tal caso los *t-test* no son adecuados, por lo que mejor emplear los *U-test* (Bootstrapping también podría ser una alternativa).
- Si el tamaño de la muestra es tal que no se puede determinar con seguridad si las poblaciones de las que proceden se distribuyen de forma normal, y no se dispone de información que pueda orientar sobre la naturaleza de las poblaciones de origen (estudios anteriores, que sea un tipo de variable que se sabe que se distribuye casi siempre de forma normal...) entonces puede ser más apropiado el *U-test* ya que no requiere asumir la normalidad de las poblaciones.

El test *U de Mann-Withney* se caracteriza por:

Comparar si dos poblaciones están equidistribuidas. Si las dos muestras proceden de la misma población, al juntar todas las observaciones y ordenarlas de menor a mayor, cabría esperar que las observaciones de una y otra muestra estuviesen intercaladas aleatoriamente. Por lo contrario, si una de las muestras pertenece a una población con valores mayores o menores que la otra población, al ordenar las observaciones, estas tenderán a agruparse de modo que las de una muestra queden por encima de las de la otra.

```
require(ggplot2)
datos <- data.frame(muestra = c(rep("A", 10), rep("B", 10)), valor = rnorm(n = 20,
  mean = 10, sd = 5), cordenada_y = rep(0, 20))

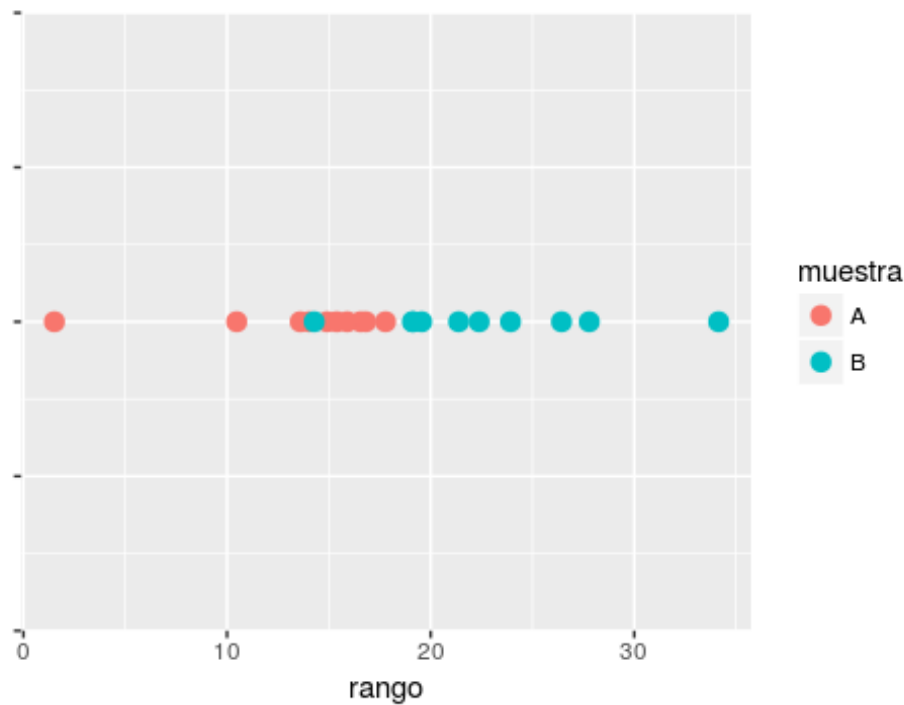
ggplot(data = datos, aes(x = valor, y = cordenada_y)) +
  geom_point(aes(colour = muestra), size = 3) + ylab("") +
  xlab("rango") +
  theme(axis.text.y = element_blank()) +
  ggtitle("Muestras procedentes de la misma población")
```



```
datos <- data.frame(muestra = c(rep("A", 10), rep("B", 10)), valor = c(rnorm(n = 10,
  mean = 10, sd = 5), rnorm(n = 10, mean = 20, sd = 5)), cordenada_y = rep(0,
  20))

ggplot(data = datos, aes(x = valor, y = cordenada_y)) +
  geom_point(aes(colour = muestra), size = 3) + ylab("") +
  xlab("rango") +
  theme(axis.text.y = element_blank()) +
  ggtitle("Muestras procedentes de distintas poblaciones")
```

Muestras procedentes de distintas poblaciones



El test *U de Mann-Withney* considera como hipótesis nula que: la probabilidad de que una observación de la población X supere a una observación de la población Y es igual a la probabilidad de que una observación de la población Y exceda a una de la población X. Es decir, que los valores de una población no tienden a ser mayores que los de otra.

$$P(X > Y) = P(Y > X)$$

Aunque la hipótesis nula descrita previamente es la descripción exacta de lo que mide el U-test, es frecuente encontrar descrito que lo que compara el U-test son las medianas. Esto no es del todo exacto pero se puede aceptar para facilitar la comprensión de los resultados.

El U-test trabaja sobre rangos de orden, es decir, utiliza las posiciones que ocupan los datos una vez ordenados. Por lo tanto, solo es aplicable a variables cuyos valores se pueden ordenar.

Al igual que ocurre con la mayoría de test no paramétricos, el *U-test* es menos potente que el *t-test* (tienen menos probabilidad de rechazar la H_0 cuando realmente es falsa) ya que ignoran valores extremos (en el caso de los t-test al trabajar con medias si los tienen en cuenta).

Esto a su vez hace que los U-test sean una prueba más robusta que los t-test. Parece ser que la pérdida de potencia es del 5%.

Condiciones necesarias del U-test

- Los datos tienen que ser independientes.
- Los datos tienen que ser ordinales, se tienen que poder ordenar de menor a mayor o viceversa.
- No es necesario asumir que las muestras se distribuyen de forma normal o que proceden de poblaciones normales. Pero sea cual sea el tipo de distribución, ambas muestras han de tener el mismo tipo.
- Requiere homogeneidad de varianza entre muestras (homocedasticidad).
- A pesar de considerarse el equivalente no paramétrico del t-test, el U-test trabaja con medianas, no con medias.
- Preferible al t-test cuando hay valores atípicos, no hay normalidad de los datos o el tamaño de las muestras es pequeño.

Ejemplo

Supóngase que se dispone de dos muestras, de las que no se conoce el tipo de distribución de las poblaciones de las que proceden y cuyo tamaño es demasiado pequeño para determinar si siguen una distribución normal. ¿Existe una diferencia significativa? *Se emplea un ejemplo con muestras pequeñas para poder ilustrar fácilmente los pasos, no significa que con muestras tan pequeñas el U-test sea muy preciso.*

```
muestraX <- c( 1.1, 3.4, 4.3, 2.1, 7.0 , 2.5 )  
muestraY <- c( 7.0, 8.0, 3.0, 5.0, 6.2 , 4.4 )
```

1.Hipotesis:

- H_0 : La probabilidad de que una observación de la población X sea mayor que una observación de la población Y es igual que la probabilidad de que una observación de la población Y sea mayor que una observación de la población X, $P(X > Y) = P(Y > X)$. Esto puede simplificarse "aceptando una serie de asunciones" que las medianas son iguales. $\text{Mediana}(X) = \text{Mediana}(Y)$.
- H_a : La probabilidad de que una observación de la población X sea mayor que una observación de la población Y no es igual que la probabilidad de que una observación de la población Y sea mayor que una observación de la población X, $P(X > Y) \neq P(Y > X)$. Las medianas de las dos poblaciones no son iguales.

2.Análisis de homocedasticidad

Para poder aplicar el U-test se requiere que la varianza sea igual en los dos grupos. Los test más recomendados para analizar homocedasticidad en estos casos son el test de *Levene* o el test de *Fligner-Killeen*, ambos trabajan con la mediana por lo que son menos sensibles a la falta de normalidad. Si se está empleando un U-test suele ser porque los datos no se distribuyen de forma normal.

```
fligner.test(x = list(muestraX,muestraY))
```

```
##  
## Fligner-Killeen test of homogeneity of variances  
##  
## data: list(muestraX, muestraY)  
## Fligner-Killeen:med chi-squared = 0.07201, df = 1, p-value =  
## 0.7884
```

No hay evidencias en contra de la igualdad de varianzas.

3.Ordenación de las observaciones

Se juntan todas las observaciones, se ordenan y se les asigna un número de posición, siendo la posición 1 para el valor más pequeño. En caso de haber valores repetidos, fenómeno conocido como ligadura o *ties*, se les asigna como valor de posición la media de las posiciones que ocupan. En este caso, el valor "7.0" está repetido y en lugar de asignarles "10" y "11" se les asigna "10.5" a cada uno.

```

muestraTotal <- sort(c(muestraX, muestraY)) # unión de las muestras y ordenación
rangosMuestra <- rank(muestraTotal)
# La función rank() de R calcula las posiciones automáticamente,
# solucionando las ligaduras en caso de que las haya.
muestraTotal <- cbind(muestraTotal, rangosMuestra)
muestraTotal

```

```

##      muestraTotal rangosMuestra
## [1,]          1.1           1.0
## [2,]          2.1           2.0
## [3,]          2.5           3.0
## [4,]          3.0           4.0
## [5,]          3.4           5.0
## [6,]          4.3           6.0
## [7,]          4.4           7.0
## [8,]          5.0           8.0
## [9,]          6.2           9.0
## [10,]         7.0          10.5
## [11,]         7.0          10.5
## [12,]         8.0          12.0

```

4.Cálculo del estadístico U

$U = \min(U_1, U_2)$ siendo U_1 y U_2 los valores estadísticos de U Mann-Whitney.

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

n_1 = tamaño de la muestra del grupo 1.

n_2 = tamaño de la muestra del grupo 2.

R_1 = sumatorio de los rangos del grupo 1.

R_2 = sumatorio de los rangos del grupo 2.

```

R1 <- sum(1,5,6,2,10.5,3)
R2 <- sum(10.5,12,4,8,9,7)
n1 <- length( muestraX )
n2 <- length( muestraY )
U1 <- (n1 * n2) + (n1*(n1 + 1) / 2) - R1
U2 <- (n1 * n2) + (n2*(n2 + 1) / 2) - R2
U <- min(c(U1,U2))
U

```

```
## [1] 6.5
```

5.Cálculo de p-value

Una vez ha obtenido el valor del estadístico U se puede calcular cual es la probabilidad de que adquiriera un valor igual o más extremo que el observado.

Si el tamaño de las muestras es inferior a 10, se compara el valor obtenido de U con los valores de una tabla U-test. Si el U es menor que el valor correspondiente en la tabla, la diferencia es significativa. En este caso el valor de la tabla para $\alpha = 0.05$, $n_1 = 6$ y $n_2 = 6$ es de 5. Dado que U obtenido es >5 no se considera significativa la diferencia. Si $n_1 > 10$ y $n_2 > 10$ se puede asumir que U se distribuye de forma aproximadamente Normal, rechazando H_0 si Z calculado es mayor que el valor de Z para el α elegido (esta es la que utiliza la función de R).

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1)$$

6.Cálculo del tamaño del efecto (side effect)

Como en todo test estadístico, no solo hay que indicar el *p-value* sino también el tamaño del efecto que se ha observado, ya que esto permitirá saber si además de significativa la diferencia es práctica. El tamaño del efecto para un U-test se calcula a partir del Z-factor (es necesario haber asumido que U se distribuye \sim Normal)

$$r = \left| \frac{Z}{\sqrt{N}} \right|, N \text{ es el número total de observaciones, } n_1 + n_2$$

La clasificación del tamaño más utilizada es:

0.1 = pequeño

0.3 = mediano

0.5 o mayor = grande

Solución directa con R

R contiene una función llamada `wilcox.test()` que realiza un U-test entre dos muestras cuando se indica que *paired* = *False* y además genera el intervalo de confianza para la diferencia de medianas.

```
wilcox.test(x = muestraX, y = muestraY, alternative = "two.sided", mu = 0, paired = FALSE, conf.int = 0.95)
```

```
## Warning in wilcox.test.default(x = muestraX, y = muestraY, alternative =  
## "two.sided", : cannot compute exact p-value with ties
```

```
## Warning in wilcox.test.default(x = muestraX, y = muestraY, alternative =  
## "two.sided", : cannot compute exact confidence intervals with ties
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: muestraX and muestraY
```

```
## W = 6.5, p-value = 0.07765
```

```
## alternative hypothesis: true location shift is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -4.9000274 0.4000139
```

```
## sample estimates:
```

```
## difference in location
```

```
## -2.390035
```

En la salida devuelta por R el valor W es el valor U.

Cuando hay ligaduras o *ties*, `wilcox.test()` no es capaz de calcular el *p-value* exacto por lo que devuelve un *p-value* aproximado asumiendo que U se distribuye de forma \sim Normal. En estos casos, o cuando los tamaños muestrales son mayores de 20 y se quiere la aproximación por

la Normal, es más recomendable emplear la función `wilcox_test()` del paquete *coin* que devuelve el valor exacto de *p-value* en lugar de una aproximación.

```
require(coin)
# La función wilcox.test() del paquete coin requiere pasarle los argumentos
# en forma de función (~), por lo que los datos tienen que estar almacenados
# en forma de data frame.
datos <- data.frame(grupo = rep(c("A", "B"), c(6, 6)), valores = c(muestraX,
  muestraY))
wilcox_test(valores ~ grupo, data = datos, distribution = "exact", conf.int = 0.95)
```

```
##
## Exact Wilcoxon-Mann-Whitney Test
##
## data:  valores by grupo (A, B)
## Z = -1.8447, p-value = 0.06926
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
##  -4.9  0.4
## sample estimates:
## difference in location
##                -2.4
```

```
# esta versión del test nos indica el valor de Z, por lo que podemos
# calcular fácilmente el tamaño del efecto
```

```
tamaño_efecto <- 1.845/sqrt(n1 + n2)
tamaño_efecto
```

```
## [1] 0.5326056
```

Resultado

La diferencia entre medianas de los dos grupos no difiere de forma significativa (*p-value* = 0.06926), el intervalo de confianza del 95% es (-4.9 , 0.4). El tamaño del efecto observado es grande (0.53) pero no significativo.

Test de Kolmogorov-Smirnov

Se trata de otro test no paramétrico para datos independientes que puede ser alternativa al U-test.