**Prueba U de Mann-Whitney**

***Descripción, condiciones y casos de uso***

Test no paramétrico que se utiliza para determinar si en dos muestras independientes, una de ellas tiende a tener un valor mayor que la otra. Puede utilizarse para determinar la asociación (“correlación”) entre una variable categórica con dos categorías y una variable continua. Para poder realizar un análisis U de Mann-Whitney, se deben cumplir las siguientes condiciones:

1. La variable dependiente debe ser ordinal o continua.
2. La variable independiente debe ser categórica y tener dos grupos.
3. Las dos observaciones deben ser independientes entre sí.

En un entorno real, a veces no se cumplen todas las condiciones. En concreto, la condición 3 puede ser más difícil de garantizar. La 1 y la 2 son obligatorias.

El resultado de la prueba es un número (*U*), que es una medida de si las muestras de una de las variables suelen quedar por encima de las otras, y un nivel de significación estadística (p-valor).

***Definición matemática***

Sean y las muestras bajo estudio. Se define el estadístico(prueba U de Mann-Whitney) como:

donde

y

En las expresiones anteriores, y corresponden a la suma de los rangos de cada una de las dos muestras, mientras que y son los tamaños de ambas. Para muestras grandes, los empates en la asignación de los rangos se pueden corregir.

***Significación estadística***

Para el cálculo del p-valor (), la hipótesis nula () es que las dos distribuciones son iguales. Obtendremos significación estadística para un determinado nivel de confianza () cuando seamos capaces de rechazar . En tal caso, podremos asegurar con significación estadística que y siguen distribuciones estadísticas diferentes.

***Aplicación a la relevancia en parquímetros***

1. Definición de variables:

- Variable independiente: Franja horaria (Sí/No). Por ejemplo, 9h-10h

- Variable dependiente: Porcentaje de ocupación

2. Hipótesis nula: estar en una franja horaria o no tiene la misma distribución estadística para el porcentaje de ocupación (no es relevante haber estado en esa franja horaria (ejemplo: 9h-10h) en el parquímetro).

3. Distribución nula: probabilidad de que sea irrelevante estar en el parquímetro durante una franja horaria en concreto (p)

4. Significación estadística (p<Alpha).

***Cálculo en R***

wilcox.test {stats}