

Estadística No Paramétrica

Clase 7 (Parte 2): Test U de Mann and Whitney

Joaquin Cavieres G.

Ingeniería en Estadística

Facultad de Ciencias, Universidad de Valparaíso



Test U de Mann and Whitney

Consideremos a $N = n + m$, en donde n = número de observaciones en X y m = número de observaciones en Y. Llamaremos a $R(X_i)$ al rango de X y $R(Y_i)$ al rango de Y en la muestral total de datos combinados. Por tanto:

- Si hay empate se asignan a los valores empatados el promedio de los rangos que le corresponderían

Así, definimos los estadísticos:

$$U_1 = \sum_{i=1}^m R(Y_i); \quad U_2 = \sum_{i=1}^n R(X_i)$$

lo que lleva finalmente a:

$$U_1 + U_2 = \frac{N(N+1)}{2}$$

Test U de Mann and Whitney

Por analogía Mann and Whitney consideraron el siguiente estadístico:

$$W = \sum_{i,j} s(Y_i - X_j) \quad (1)$$

donde las variables que participan en la sumatoria no son independientes. Por tanto, si no hay empates (datos repetidos)

$$U = W + \frac{m(m+1)}{2} \quad (2)$$

Test U de Mann and Whitney

Hipótesis

$$H_0 : \Delta = 0$$

Entonces,

a)

$$P(R(Y_i) = k) = 1/k, \quad 1 \leq k \leq N$$
$$P(R(Y_i) = k, R(Y_j)) = \begin{cases} \frac{1}{N(N-1)} & \text{si } k \neq l, i \neq j \\ 0 & \text{en caso contrario} \end{cases}$$

b)

$$\mathbb{E}(R(Y_i)) = \frac{N+1}{2}, \quad \text{VAR}(R(Y_i)) = \frac{N^2-1}{12}$$

Test U de Mann and Whitney

Si existen valores repetidos (empates), entonces la esperanza de U no cambia pero sí su varianza. Por tanto, si s es el número de grupos con valores repetidos y d_j es el número de valores repetidos en el grupo j , tenemos:

$$\text{VAR}(U/d_1, \dots, d_s) = mn \frac{N+1}{12} - \frac{mn \sum_{i=1}^s (d_i^3 - d_i)}{12N(N-1)} \quad (3)$$

así la varianza de U es:

$$\text{VAR}(\hat{U}) = \frac{mn}{N(N-1)} \sum_{i=1}^N R_i^2 - \frac{mn(N+1)^2}{4(N-1)} \quad (4)$$

Test U de Mann and Whitney

Ejemplo 2

Dentro de un curso de la carrera de Ingeniería Estadística hay 48 alumnos, de los cuales 12 viven en el campo y 36 viven en la ciudad. Dentro de sus ramos optativos (educación física) se les tomó un test para medir su condición asignandoles a cada uno una puntuación. Los resultados son los siguientes:

Campo (X_i)		Ciudad (Y_i)					
14.8	10.6	12.7	16.9	7.6	2.4	6.2	9.9
7.3	12.5	14.2	7.9	11.3	6.4	6.1	10.6
5.6	12.9	12.6	16.0	8.3	9.1	15.3	14.8
6.3	16.1	2.1	10.6	6.7	6.7	10.6	5.0
9.0	11.4	17.7	5.6	3.6	18.6	1.8	2.6
4.2	2.7	11.8	5.6	1.0	3.2	5.9	4.0

Se desea testear: H_0 : Los alumnos (campo y ciudad) tienen la misma condición física.

Desarrollo

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

Luego de ordenar las puntuaciones en los test para todos los alumnos (campo y ciudad) y su correspondiente rango, calculamos el estadístico:

$$U_1 = \sum_{i=1}^n R(X_i) =$$

Test U de Mann and Whitney

Desarrollo

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

Luego de ordenar las puntuaciones en los test para todos los alumnos (campo y ciudad) y su correspondiente rango, calculamos el estadístico:

$$U_1 = \sum_{i=1}^n R(X_i) = 321.$$

¿Como deberíamos proceder para una aproximación Normal?

Test U de Mann and Whitney

Desarrollo

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

Luego de ordenar las puntuaciones en los test para todos los alumnos (campo y ciudad) y su correspondiente rango, calculamos el estadístico:

$$U_1 = \sum_{i=1}^n R(X_i) = 321.$$

¿Como deberíamos proceder para una aproximación Normal? (TAREA)

Test U de Mann and Whitney

Desarrollo

$$H_0 : \Delta = 0$$

$$H_1 : \Delta \neq 0$$

Vamos a rechazar H_0 con nivel de significancia $\alpha = 0.05$ si el estadístico es mayor a 1.65, entonces:

Calculos en R

Ver ejemplo...

Test U de Mann and Whitney

Observaciones importantes

Este test se usa como alternativa en ocasiones al t-test cuando las muestras no provienen de una distribución Normal. Hay que tener en cuenta que:

- Ambos test permiten realizar inferencia sobre la diferencia entre poblaciones pero las hipótesis no son las mismas.
- El t-test compara media entre los grupos, en cambio el test U es menos específico.
- Si las distribuciones entre poblaciones difieren en la localización, el test U compara medianas, para otros escenarios (poblaciones con distinta distribución, dispersión, asimetría, etc) contrasta $P(X > Y) = P(Y > X)$. En este caso el p-value no puede interpretarse directamente.

Ejemplo 3

Ver ejemplo en R