Estadística No Paramétrica

Clase 7: Test U de Mann and Whitney

Joaquin Cavieres G.

Ingeniería en Estadística

Facultad de Ciencias, Universidad de Valparaíso

Test U de Mann and Whitney

El contenido de esta segunda parte del curso se centrará sobre dos muestras aleatorias independientes.

Test U de Mann and Whitney

Los datos son $N = m + n$ observaciones de $X_1, ..., X_m$ y $Y_1, ..., Y_n$

Supuestos

- Las observaciones $X_1, ..., X_m$ son variables aleatorias desde una población 1, esto es, que todas las $X's$ son *i.i.d.* Las observaciones $Y_1, ..., Y_n$ son variables aleatorias desde una población 2, esto es, que todas las $Y's$ son *i.i.d.*
- Las $X's$ e $Y's$ son mutuamente independiente, esto significa que ademas de que las muestras sean independientes, existe independencia entre las dos muestras.
- La población 1 y la población 2 son continuas.

Test U de Mann and Whitney

Hipótesis

Se tiene una función de distribución F correspondiente a la población 1 y una función de distribución G correspondiente a la población 2. Por tanto, se plantea la siguiente hipótesis nula:

$$H_0 : F(t) = G(t) \text{ para todo } t$$

Lo anterior nos dice que la variable X y la variable Y tienen la misma función de distribución pero no se especifíca la distribución en común.

Hipótesis

Supuestos

Tal como se presenta la hipótesis parecería ser un test para diferencia entre distribuciones, es decir que la hipótesis nula sería $H_0 : F(t) = G(t) \ \forall \ t$, siendo F la distribución de la $X_1, ..., X_m$ y G la de $Y_1, ..., Y_n$. Sin embargo, aquí se trata como un test para el parámetro de posición, para lo cual supondremos que $G(t) = F(t - \Delta)$, para algún Δ.

Test U de Mann and Whitney

Entonces:

$$G(t) = F(t - \Delta) \tag{1}$$

Indica que la población 2 es la misma que la población 1 excepto que esta es "desplazada" por una cantidad Δ. Esto se conoce como un modelo de cambio de ubicación (*location-shift model.*). Este tipo de modelos también se puede escribir como:

$$Y \stackrel{d}{=} X - \Delta \tag{2}$$

donde el simbólo $\stackrel{d}{=}$ significa "tiene la misma distribución como" y el simbólo Δ es llamado "cambio de ubicación" o "desplazamiento de ubicación" o "efecto del tratamiento".

Ejemplo

Si X es un valor seleccionado aleatoriamente desde la población 1, la "población de control", e Y es un valor seleccionado aleatoriamente de la población 2, la "población de tratamiento", entonces Δ es el efecto esperado debido al tratamiento. Si es positivo, es el aumento esperado debido al tratamiento, y si es negativo, es la disminución esperada debido al tratamiento.

Lo anterior se expresa puede expresar matemáticamente como:

$$\Delta = \mathbb{E}(Y) - \mathbb{E}(X), \tag{3}$$

como las diferencias en las medias de las poblaciones.

Test U de Mann and Whitney

La hipótesis nula H_0 entonces se reduce a:

$$H_0 : \Delta = 0, \tag{4}$$

afirmando que las medias poblacionales son iguales o, de manera equivalente, que el tratamiento no tiene efecto.

Test U de Mann and Whitney

Forma de calculo

Para calcular el estadístico W debemos:

- Ordenar la muestra $N = m + n$ valores X y valores Y de menor a mayor.
- S_1 denota el rango de $Y_1, ..., S_n$ en este orden conjunto.
- W es la suma del rango asignado para los valores de Y,

$$W = \sum_{j=1}^{n} S_j \tag{5}$$

Test U de Mann and Whitney

a) Prueba de cola superior

$$H_0 : \Delta = 0$$
$$H_1 : \Delta > 0$$

entonces, con un nivel de signifcancia α, Rechazamos H_0 si $W \geq w_\alpha$, de otra manera no rechazamos H_0.

Aquí w_α es elegido para hacer que la probabilidad del error tipo 1 sea igual a α.

b) Prueba de cola inferior

$$H_0 : \Delta = 0$$
$$H_1 : \Delta < 0$$

entonces, con un nivel de signifcancia α, Rechazamos H_0 si $W \leq n(m+n+1) - w_{\alpha}$, de otra manera no rechazamos H_0.

Test U de Mann and Whitney

c) Prueba de ambas colas

$$H_0 : \Delta = 0$$
$$H_1 : \Delta \neq 0$$

entonces, con un nivel de signifcancia α, Rechazamos H_0 si $W \geq w_{\alpha/2}$ o si $W \leq n(m+n+1) - w_{\alpha/2}$, de otra manera no rechazamos H_0.

La prueba de ambas colas es simétrica con $\alpha/2$ probabilidad en cada extremo de la distribución.

Aproximación mediante n grande

La aproximación mediante un n grande esta basada en la normalidad asintótica de W correctamente estandarizada. Para esto necesitamos conocer la media y varianza cuando H_0 es verdadera. Si H_0 es verdadera entonces la media y la varianza son:

$$\mathbb{E}_0(W) = \frac{n(m+n+1)}{2} \tag{6}$$

y

$$\mathbb{VAR}_0(W) = \frac{mn(m+n+1)}{12} \tag{7}$$

Test U de Mann and Whitney

Aproximación mediante *n* grande

La versión estandarizada de W es:

$$W^* = \frac{W - \mathbb{E}_0(W)}{\mathbb{VAR}_0(W)}^{1/2} = \frac{W - (n - (m+n+1)/2)}{mn(m+n+1)/12}^{1/2} \tag{8}$$

Cuando H_0 es verdadera, W^* tiene, cuando $min(m,n) \to \infty$, una distribución $N(0,1)$.

Por lo tanto,

Test U de Mann and Whitney

Aproximación mediante n grande

La versión estandarizada de W es:

$$W^* = \frac{W - \mathbb{E}_0(W)}{\mathbb{VAR}_0(W)}^{1/2} = \frac{W - (n - (m+n+1)/2)}{mn(m+n+1)/12}^{1/2} \qquad (8)$$

Cuando H_0 es verdadera, W^* tiene, cuando $min(m,n) \to \infty$, una distribución $N(0,1)$.

Por lo tanto,

- Rechazamos H_0 si $W^* \geq z_\alpha$, de otra manera no rechazamos H_0 para la cola superior.
- Rechazamos H_0 si $W^* \leq -z_\alpha$, de otra manera no rechazamos H_0 para la cola inferior.
- Rechazamos H_0 si $|W^*| \geq z_{\alpha/2}$, de otra manera no rechazamos H_0 para la ambas colas.

Test U de Mann and Whitney

Estadístico Mann–Whitney

Los procedimientos para la cola superior, inferior y ambas colas basados en el rango puede ser realizado mediante el estadístico Mann–Whitney:

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} \phi(X_i, Y_j), \tag{9}$$

donde

$$\phi(X_i, Y_j) = \begin{cases} 1 & \text{si } X_i < Y_j \\ 0 & \text{si en otro caso} \end{cases} \tag{10}$$

Test U de Mann and Whitney

El estadístico U "cuenta" el número de X predecesores de Y. Es posible mostrar que:

$$W = U + \frac{n(n+1)}{2} \tag{11}$$

Por esto el test W y el test U son equivalentes.

Test U de Mann and Whitney

Ejemplo