Estadística III

Pruebas de 1 y 2 muestras

Alejandro López Hernández

FES Acatlán - UNAM

April 4, 2020

Índice

Las pruebas de 1 una muestra consiste en tener una población a la cual queremos comparar en dos instacias distintas, por ejemplo, cuando se aplica cierto tratamiento y se quiere medir el impacto de tal tratamiento. Asumiremos que tenemos $2n$ datos tales que $Z_i = X_i - Y_i$ son independientes, tambien asumiremos que Z_i tienen la misma distribución la cual tiene media θ, este parmetro será el de nuestro interes.

Está prueba sirve para probar la hipotesis $H_0 : \theta = 0$, para probar esa hipotesis se costruye el estadístico T^+ con los rangos de los valores de $|Z_i|$, se define el estadístico como

$$T^+ = \sum_{i=1}^{n} R_i \varphi_i$$

donde φ_i vale 1 cuando $|Z_i| > 0$ y 0 cuando $|Z_i| < 0$.

Se pueden realizar 3 pruebas:

- Prueba de la cola superior, $H_0 : \theta = 0$ vs $H_1 : \theta > 0$, H_0 se debe rechazar con una significancia de α si $T^+ \geq t_\alpha$.
- Prueba de la cola inferior, $H_0 : \theta = 0$ vs $H_1 : \theta < 0$, H_0 se debe rechazar con una significancia de α si $T^+ \leq \frac{n(n+1)}{2} - t_\alpha$.
- Prueba de dos colas, $H_0 : \theta = 0$ vs $H_1 : \theta = 0$, H_0 se debe rechazar con una significancia de α si $T^+ \geq t_{\alpha/2}$ o $T^+ \leq \frac{n(n+1)}{2} - t_{\alpha/2}$

Se puede realizar una modificación en el estadístico para poder comparlo contra los cuantiles de la normal, esto debido a que $\mathbb{E}_0(T^+) = \frac{n(n+1)}{4}$ y que $\text{Var}(T^+)_0 = \frac{n(n+1)(2n+1)}{12}$, por lo tanto podemos construir el esadístico

$$T^* = \frac{T^+ - \frac{n(n+1)}{4}}{\left(\frac{n(n+1)(2n+1)}{12}\right)^{1/2}}$$

Con T^* las pruebas quedan como:

- Prueba de la cola superior, $H_0 : \theta = 0$ vs $H_1 : \theta > 0$, H_0 se debe rechazar con una significancia de α si $T^* \geq z_\alpha$.
- Prueba de la cola inferior, $H_0 : \theta = 0$ vs $H_1 : \theta < 0$, H_0 se debe rechazar con una significancia de α si $T^* \leq -t_\alpha$.
- Prueba de dos colas, $H_0 : \theta = 0$ vs $H_1 : \theta = 0$, H_0 se debe rechazar con una significancia de α si $|T^*| \geq z_{\alpha/2}$

Con las pruebas de 2 muestras, nuestro interes será comparar un par de muestras que provienen de poblaciones con alguna caracteristica que las diferencian, y queremos investigar en si esa característica es significativa. Otra forma de verlo es con el caso particular en el que tenemos una poblacion a la cual le aplicamos cierto tratamiento, entonces queremos ver si el tratamiento tuvo un efecto *positivo* o *negativo*.

Esta prueba consiste en probar la hipotesis de que ambas muestras tienen la misma distribución, es decir $H_0 : F(t) = G(t)$, sin embargo en ningún momento se especifica cual es la distribución de F ó G, por lo tanto podemos cambiar la prueba a $H_0 : \Delta = \mathbb{E}(X) - \mathbb{E}(Y) = 0$. La forma de construir el estadístico es calcular los rangos conjuntos de X y Y, con sus $N = m + n$, si S_j denota los rangos de Y, entonces el estadístico se calcula como:

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

Se pueden realizar 3 pruebas:

- Prueba de la cola superior, $H_0 : \Delta = 0$ vs $H_1 : \Delta > 0$, H_0 se debe rechazar con una significancia de α si $W \geq \omega_\alpha$.
- Prueba de la cola inferior, $H_0 : \Delta = 0$ vs $H_1 : \Delta < 0$, H_0 se debe rechazar con una significancia de α si $W \leq n(m+n+1) - \omega_\alpha$.
- Prueba de dos colas, $H_0 : \Delta = 0$ vs $H_1 : \Delta = 0$, H_0 se debe rechazar con una significancia de α si $W \geq \omega_{\alpha/2}$ o $W \leq n(m+n+1) - \omega_{\alpha/2}$.

En el caso de que tengamos una gran cantidad de datos N, podemos calcular un estadístico alterno que se pueda comparar con una normal, eso gracias al hecho de que $\mathbb{E}_0(W) = \frac{n(m+n+1)}{2}$ y que $\text{Var}_0(W) = \frac{mn(m+n+1)}{12}$, por lo tanto podemos construir el estadístico:

$$W^* = \frac{W - (n(m+n+1)/2)}{(mn(m+n+1)/12)^{1/2}}$$

Prueba de Mann-Whitney-Wilcoxon

De esta forma las pruebas se convierten en:

- Prueba de la cola superior, $H_0 : \Delta = 0$ vs $H_1 : \Delta > 0$, H_0 se debe rechazar con una significancia de α si $W^* \geq z_\alpha$.
- Prueba de la cola inferior, $H_0 : \Delta = 0$ vs $H_1 : \Delta < 0$, H_0 se debe rechazar con una significancia de α si $W^* \leq -z_\alpha$
- Prueba de dos colas, $H_0 : \Delta = 0$ vs $H_1 : \Delta = 0$, H_0 se debe rechazar con una significancia de α si $|W^*| \geq z_{\alpha/2}$