Estadística No Paramétrica

Clase 8: Test suma de rangos Wilcoxon y Test para reacciones extremas (Moses)

Joaquin Cavieres G.

Ingeniería en Estadística

Facultad de Ciencias, Universidad de Valparaíso

Test para muestras independientes (location problem)

Recordatorio: El contenido de la segunda parte del curso se centrará sobre dos muestras aleatorias independientes.

Test para muestras independientes (location problem)

Supuestos

- $X_1, ..., X_m$ son variables aleatorias desde una población 1
- $Y_1, ..., Y_n$ son variables aleatorias desde una población 2
- Dado los puntos anteriores, $X_1, ..., X_m$ e $Y_1, ..., Y_n$ son *i.i.d.*
- $(X_1, ..., X_m)$ y el vector $(Y_1, ..., Y_n)$ son independientes.
- Función de distribución de la población 1 es $F(t)$ y la función de distribución de la población 2 es $G(t)$.

Test de suma de rangos Wilcoxon

Hipótesis nula H_0

Bajo estos supuestos, estamos interesados en evaluar si hay alguna diferencia entre las distribuciones de probabilidad para X e Y, esto es en forma general:

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

con $F(t) = Pr\{X_i \leq t\}$ y $G(t) = Pr\{Y_i \leq t\}$

Test de suma de rangos Wilcoxon

Hipótesis aleternativa H_1

$$H_1 : G(t) = F(t - \Delta)$$

Entonces bajo H_0

$$Pr\{X_i + \Delta \leq t\} = Pr\{X_i \leq t - \Delta\} = F(t - \Delta) = G(t) \Rightarrow$$
$$\Rightarrow X_1 + \Delta, ..., X_m + \Delta; Y_1, ..., Y_n : \text{i.i.d}$$

Estimador Hodges-Lehmann Este estimador está asociado con el test de suma de rangos Wilcoxon de la siguiente manera:

Desde mn diferencias $Y_j - X_i$, para un $i = 1, ..., m$ y un $j = 1, ..., n$, el estimador de Δ es:

$$\hat{\Delta} = \text{mediana}\{(Y_j - X_i)\}, i = 1, ..., m; j = 1, .., n$$

Estimador Hodges-Lehmann

$$\hat{\Delta} = \text{mediana}\{(Y_j - X_i)\}, i = 1, ..., m; j = 1, .., n$$

Considere a $U^1 \leq ... \leq U^{mn}$ como los valores ordenados de $Y_j - X_i$ entonces, si mn es impar (digamos $mn = 2k + 1$), $k = (mn - 1)/2$ y:

$$\hat{\Delta} = U^{k+1},$$

es el valor que ocupa la posición $k + 1$ en la lista ordenada de $Y - X$.

Estimador Hodges-Lehmann

$$\hat{\Delta} = \text{mediana}\{(Y_j - X_i)\}, i = 1, ..., m; j = 1, .., n$$

Si mn es par (digamos $mn = 2k$), $k = mn/2$ y:

$$\hat{\Delta} = \frac{U^k + U^{k+1}}{2}$$

Esto es, $\hat{\Delta}$ es el promedio de dos veces la diferencia $Y - X$ que ocupa la posición k y $k + 1$ en la lista ordenada de mn diferencias.

Test de suma de rangos Wilcoxon

Estimador Hodges-Lehmann

$$\hat{\Delta} = \text{mediana}\{(Y_j - X_i)\}, i = 1, ..., m; j = 1, .., n$$

Comentarios:

- Un razónable estimador de Δ es la cantidad $\hat{\Delta}$ que debería ser sustraída desde cada Y_j para que el valor de W (cuando es aplicado a la muestra ordenada) es $n(m + n + 1)/2$.
- El estimador $\hat{\Delta}$ es menos sensible a la diferencia de $\bar{Y} - \bar{X}$ en la teoría clásica.

Test de suma de rangos Wilcoxon

Intervalos de confianza usando la suma de rangos propuesta por Wilcoxon.

Derivación

Considerar a Δ_0 tal que el test para ambas colas:

$$H_0 : \Delta = \Delta_0 \qquad \text{v/s} \qquad H_1 : \Delta \neq \Delta_0$$

Intervalo de confianza

$$C_\alpha = \frac{n(2m+n+1)}{2} + 1 - w_{\alpha/2}$$

para $U_1 < U_2 <, ..., < U_{mn}$ (Estadísticos ordenados de $\{(Y_j - X_i) i = 1, ..., m; j = 1, ..., n\}$), entonces:

$$\Delta_L = U_{C_\alpha}, \;\; \Delta_U = U_{mn+1-C_\alpha} \Rightarrow \Pr_{\Delta}\{\Delta_L < \Delta < \Delta_U\} = 1 - \alpha \;\; \forall \Delta$$

Test de suma de rangos Wilcoxon

$$H_0 : \Delta = \Delta_0 \qquad \text{v/s} \qquad H_1 : \Delta \neq \Delta_0$$

Intervalo de confianza

$$\underset{\Delta}{Pr}\{\Delta_L < \Delta < \Delta_U\} = 1 - \alpha$$

El intervalo de confianza puede encontrarse directamente en `R` con la función `wilcox.test()`

Test de suma de rangos Wilcoxon (Moses)

Aproximación mediante n grande

Para un número grande de m y n, C_α puede ser aproximado mediante:

$$C_\alpha = \frac{mn}{2} - z_{\alpha/2}\left\{\frac{mn(m+n+1)}{12}\right\}^{1/2} \tag{1}$$

Test de reacciones extremas de Moses

A veces, en algunos experimentos, diferentes grupos reaccionan en forma distinta a tratamientos específicos. Por ejemplo, en depresiones económicas algunas personas se vuelvan extremadamente reaccionarias y otras extremadamente polarizadas por algún pensamiento político. O podriamos esperar que el cambio climático cree provoque temperaturas extremas en algunos sectores y lluvias torrenciales en otros.

La prueba de Moses está diseñada específicamente para su uso con datos (medidos al menos en una escala ordinal) recopilados para probar tales hipótesis. Debe usarse cuando se espera que la condición experimental afecte a algunos sujetos de una manera y a otros de manera opuesta

Test de reacciones extremas de Moses (The Moses extreme reaction test)

La test de reacciones de Moses está diseñado específicamente para trabajar con datos medidos en una escala ordinal. Debe usarse cuando se espera que la condición experimental afecte a algunos sujetos de una manera y a otros de manera opuesta.

Test de reacciones extremas de Moses (The Moses extreme reaction test)

Test para evaluar dos muestras aleatorias en escala continua:

$$H_0 := \text{Valores extremos son igualmente probables en ambas poblaciones}$$
$$H_1 : \text{Valores extremos no son igualmente probables en ambas poblaciones}$$

Test de reacciones extremas de Moses

Ejemplos

Ver ejemplos en `R`