Estadística III

Pruebas de independencia

Alejandro López Hernández

FES Acatlán - UNAM

March 27, 2020

Índice

El problema que intentaremos resolver es cuando tenemos una muestra bivariada y queremos saber la relación que existe entre las dos variables aleatorias, en particular la indepenecia. Nuestro supuesto es que tenemos una muestra bivariada independiente de n datos, del tipo $(X_1, Y_2), \ldots, (X_n, Y_n)$. La hipótesis que queremos probar es de la forma

$$H_0 : [F_{XY}(x, y) = F_X(x)F_Y(y) \text{ para cualquier par } (x, y)]$$

La prueba de Kendall se basa en la cantidad $\tau = 2\mathbb{P}((Y_2 - Y_1)(X_2 - X_1) > 1) - 1$, está cantidad se propone debido a que si X fuera independiente de Y, τ deberia de ser 0, por lo tanto nuestra prueba de hipotesis la probaremos buscando valores pequeños de τ, sin embargo debemos probar todas las combinaciones de pares entre las observaciones.

Para el cálculo del estadístico, utilizamos la siguiente función

$$Q((a,b),(c,d)) = \begin{cases} 1 & \text{si } (d-b)(c-a) > 0 \\ -1 & \text{si } (d-b)(c-a) < 0 \end{cases}$$

El estadístico de Kendall se define como:

$$K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q((X_i, Y_i),(X_j, Y_j))$$

Para calcular la distribución de K, se puede aproximar con la distrubición normal, para eso utilizamos el hecho de que $\mathbb{E}(K) = 0$ y $\text{Var}(K) = \frac{n(n-1)(2n+5)}{18}$, con ello podemos modificar el estadístico como:

$$K^* = \frac{K}{(n(n-1)(2n+5)/18)^{1/2}}$$

La prueba de Spearman, no resulta ser tan intuitiva, se define como la correlación de los rangos de los datos, es decir:

$$r_s = \frac{12 \sum_{i=1}^{n} [R_i - \frac{n+1}{2}][S_i - \frac{n+1}{2}]}{n(n^2 - 1)}$$

de igual forma se busca que r_s sea una cantidad baja cuando la hipótesis sea cierta.

De igual forma se puede aproximar cuando se tiene una gran cantidad de datos por una normal, tenemos que $\mathbb{E}(r_s) = 0$ y $\text{Var}(r_s) = \frac{1}{n-1}$, por lo tanto podemos modificar el estadístico como:

$$r_s^* = \sqrt{n-1}r_s$$

Y las regiones de rechazo las podemos poner en terminos de la distribución normal.