Correlación de Rangos

Se utiliza con datos apareados para probar una asociación entre dos variables.

Definición

La **prueba de correlación de rangos** (o **prueba de correlación de rangos de Spearman**) es una prueba no paramétrica que utiliza rangos de datos muestrales consistentes en datos apareados. Se utiliza para probar una asociación entre dos variables, por lo que las hipótesis nula y alternativa son las siguientes (donde ρ_s denota el coeficiente de correlación de rangos de la población completa):

$H0$: $\rho_s = 0$ (*No* existe correlación entre las dos variables).

$H1$: $\rho_s \neq 0$ (Existe una correlación entre las dos variables).

Ventajas: La correlación de rangos tiene las siguientes ventajas sobre los métodos paramétricos:

1. El método no paramétrico de correlación de rangos puede utilizarse en una variedad

más amplia de circunstancias que el método paramétrico de correlación lineal.

Con la correlación de rangos podemos analizar datos apareados que sean rangos o puedan convertirse en rangos. Por ejemplo, si dos jueces califican a 30 gimnastas, podemos utilizar la correlación de rangos, pero no la correlación lineal. A diferencia de los métodos paramétricos, el método de correlación de rangos *no* requiere una distribución normal de cualquier población.

2. La correlación de rangos puede utilizarse para detectar algunas relaciones (no todas) que no son lineales.

Correlación de Rangos

Requisitos

1. Los datos muestrales apareados se seleccionaron aleatoriamente.

2. A diferencia de los métodos paramétricos, no existe el requisito de que los datos muestrales apareados tengan una distribución normal bivariada. No existe el requisito de una distribución normal para cualquier población.

Notación

r_s =coeficiente de correlación de rangos para datos muestrales apareados (r_s es un estadístico muestral).

ρ_s =coeficiente de correlación de rangos para todos los datos poblacionales (ρ_s es un parámetro poblacional).

n=número de pares de datos muestrales.

d=diferencia entre los rangos de los dos valores dentro de un par.

Estadístico de prueba

Sin empates: Después de convertir los datos de cada muestra a rangos, si no existen empates entre los rangos para la primera variable y no existen empates entre los rangos para la segunda variable, el valor exacto del estadístico de prueba puede calcularse utilizando esta fórmula:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

Correlación de Rangos

Empates: Después de convertir los datos de cada muestra a rangos, si cualquier variable tiene empates entre sus rangos, el valor exacto del estadístico de prueba r_s puede calcularse utilizando la fórmula con los rangos:

$$r_s = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}}$$

Valores críticos

1. Si $n \leq 30$, los valores críticos se encuentran en la tabla A-9. (Dar clic para poder visualizar la tabla)

2. Si $n > 30$, los valores críticos de r_s se calculan utilizando la siguiente fórmula:

$$r_s = \frac{\pm z}{\sqrt{n-1}}$$

(valores críticos cuando *n* > 30)

Donde el valor de *z* corresponde al nivel de significancia. (Por ejemplo, si α =0.05, *z*=1.96).

EJEMPLO

Se incluyen las clasificaciones de universidades por parte de estudiantes y de la revista *U.S. News and World Report*. Esta clasificación está incluida en la tabla 5, que también incluye las diferencias *d* y los cuadrados de las diferencias d^2.

Correlación de Rangos

Tabla 1. Universidades clasificadas por estudiantes y por U.S. News

Universidad	Clasificación según la preferencia de los estudiantes	Clasificación según la revista *U.S. News and World Report*	Diferencia d	d^2
Harvard	1	1	0	0
Yale	2	2	0	0
Cal. Inst. of Tech.	3	5	2	4
M.I.T.	4	4	0	0
Brown	5	7	2	4
Columbia	6	6	0	0
U. de Penn.	7	3	4	16
Notre Dame	8	8	0	0
			Total	24

Calcule el valor del coeficiente de correlación de rangos y utilícelo para determinar si existe una correlación entre la clasificación de los estudiantes y la clasificación de la revista. Use un nivel de significancia de 0.05.

Solución

El único requisito es que los datos muestrales apareados sean elegidos al azar. Las universidades incluidas se eligieron al azar entre aquellas que estaban disponibles, de manera que procedemos con la prueba. El coeficiente de correlación lineal *r* no debe utilizarse puesto que requiere de distribuciones normales, y los datos consisten en rangos que no están distribuidos normalmente. En vez de ello, utilizamos el coeficiente de correlación de rangos para probar una relación entre los rangos de los estudiantes y de la revista.

Las hipótesis nula y alternativa son las siguientes:

$$H_0: \rho_s = 0$$

$$H_a: \rho_s \neq 0$$

Siguiendo el procedimiento, los datos están en forma de rangos y ninguna de las dos variables tiene empates entre los rangos, por lo que el valor exacto del estadístico de prueba puede calcularse normalmente. Utilizamos n= 8 (para 8 pares de datos) y $\sum d^2 = 24$ (como se indica en la tabla de 5) para obtener:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6(24)}{8(8^2-1)} = 1 - \frac{144}{504} = 0.714$$

Ahora nos remitimos a la tabla de A-9 para determinar que los valores críticos son ±0.738 (con α =0.05 y n =8). Puesto que el estadístico de prueba $r_s = 0.714$ no excede al valor crítico de 0.738, no rechazamos la hipótesis nula. No existe suficiente evidencia para sustentar una aseveración de correlación entre la clasificación de los estudiantes y la clasificación de la revista. Parece que en lo que se refiere a la clasificación de universidades, los estudiantes y la revista no coinciden. (Si coincidieran, habría una correlación significativa, pero no la hay).

Referencia:

Triola, M., (2013). Estadística. Decimoprimera edición. Pearson educación. México