Prueba de Kruskal -Wallis

En esta lección se describe la *prueba de Kruskal-Wallis*, que utiliza rangos de datos de tres o más muestras independientes para probar la hipótesis nula de que las muestras provienen de poblaciones con medianas iguales.

Definición

La **prueba de Kruskal-Wallis** (también llamada la prueba H) es una prueba no paramétrica que utiliza rangos de datos muestrales de tres o más poblaciones independientes. Se utiliza para probar la hipótesis nula de que las muestras independientes provienen de poblaciones con medianas iguales; la hipótesis alternativa es la aseveración de que las poblaciones tienen medianas que no son iguales.

H0: Las muestras provienen de poblaciones con medianas iguales.

H1: Las muestras provienen de poblaciones con medianas que no son iguales.

Para aplicar la prueba de Kruskal-Wallis, calculamos el *estadístico de prueba H, el cual tiene una distribución que puede aproximarse por medio la distribución chi cuadrada, siempre y cuando cada muestra tenga al menos cinco observaciones*.

Requisitos

1. Tenemos al menos tres muestras independientes, las cuales se seleccionan al azar.

2. Cada muestra tiene al menos cinco observaciones. (Si las muestras tienen menos de cinco observaciones, remítase a tablas especiales de valores críticos, como las *CRC Standard Probability and Statistics Tables and Formulae*, publicadas por CRC Press).

3. *No* existe el requisito de que las poblaciones tengan una distribución normal o alguna otra distribución particular.

Notación

N=número total de observaciones en todas las muestras combinadas.

K= número de muestras.

$R1$= suma de los rangos de la muestra 1.

$n1$= número de observaciones de la muestra 1.

Para la muestra 2, la suma de los rangos es $R2$ y el número de observaciones es $n2$, y se utiliza una notación similar para las otras muestras.

Estadístico de prueba

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

Valores críticos

1. La prueba es de cola derecha.

2. gl= k-1. (Puesto que el estadístico de prueba H puede aproximarse por medio de una distribución chi cuadrada, utilice la tabla A-4 (Dar clic para visualizar la tabla) con k-1 grados de libertad, donde k es el número de muestras diferentes).

Prueba de Kruskal -Wallis

Procedimiento para calcular el valor del estadístico de prueba H

1. Combine temporalmente todas las muestras en una muestra grande y asigne un rango a cada valor muestral. (Ordene los valores del menor al mayor, y en caso de empates, asigne a cada observación la media de los rangos implicados).

2. En cada muestra, calcule la suma de los rangos y calcule el tamaño muestral.

3. Calcule *H* utilizando los resultados del paso 2, con la notación y el estadístico de prueba descritos anteriormente.

El estadístico de prueba *H* es básicamente una medida de la varianza de las sumas de rangos $R1, R2, \ldots, Rk$. Si los rangos están distribuidos de forma equitativa entre los grupos muestrales, entonces *H* debe ser un número relativamente pequeño. Si las muestras son muy diferentes, entonces los rangos serán excesivamente bajos en algunos grupos y altos en otros, con el efecto neto de que *H* será grande. En consecuencia, solo los valores grandes de *H* nos llevan al rechazo de la hipótesis nula de que las muestras provienen de poblaciones idénticas. *La prueba de Kruskal-Wallis es, por lo tanto, una prueba de cola derecha.*

EJEMPLO

La tabla 4 lista los pesos (en kg) de álamos que recibieron tratamientos diferentes.

Tabla 1. Pesos en Kilogramos de álamos

Tratamiento			
Ninguno	Fertilizante	Riego	Fertilizante y riego
0.15 (8)	1.34 (18)	0.23 (12)	2.03 (19)
0.02 (1.5)	0.14 (7)	0.04 (3)	0.27 (13)
0.16 (9.5)	0.02 (1.5)	0.34 (14)	0.92 (16)
0.37 (15)	0.08 (5.5)	0.16 (9.5)	1.07 (17)
0.22 (11)	0.08 (5.5)	0.05 (4)	2.38 (20)
$n1=5$	$n2=5$	$n3=5$	$n4=5$
$R1=45$	$R2=37.5$	$R3=42.5$	$R4=85$

Use la prueba de Kruskal-Wallis para probar la hipótesis nula de que las cuatro muestras provienen de poblaciones con medianas iguales.

Para determinar el valor del estadístico de prueba H, primero tenemos que ordenar en rangos todos los datos. Comenzamos con los valores más bajos de 0.02 y 0.02. Como hay un empate entre los valores correspondientes a los rangos 1 y 2, asignamos el rango medio de 1.5 a cada uno de los valores empatados. En la tabla 4 los rangos aparecen entre paréntesis, junto al peso original del árbol. Después calculamos el tamaño muestral n y la suma de rangos R para cada muestra, los cuales se presentan en la parte inferior de la tabla 4. Puesto que el número total de observaciones es 20, tenemos N=20. Ahora podemos evaluar el estadístico de prueba de la siguiente manera:

Prueba de Kruskal -Wallis

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

$$H = \frac{12}{20(20+1)}\left(\frac{45^2}{5} + \frac{37.5^2}{5} + \frac{42.5^2}{5} + \frac{85^2}{5}\right) - 3(20+1)$$

$$H = 8.214$$

Puesto que cada muestra tiene al menos cinco observaciones, la distribución de H es aproximadamente una distribución chi cuadrada con k- 1 grados de libertad. El número de muestras es k= 4, entonces tenemos 4-1= 3 grados de libertad. Remítase a la tabla A-4 para encontrar el valor crítico de 7.815, que corresponde a 3 grados de libertad y a un nivel de significancia de 0.05 (con un área de 0.05 en la cola derecha). El estadístico de prueba H=8.214 está en la región crítica acotada por 7.815; por lo tanto, rechazamos la hipótesis nula de medianas iguales.

INTERPRETACIÓN

Existe suficiente evidencia para rechazar la aseveración de que las poblaciones de los pesos de álamos con los cuatro tratamientos tienen medianas iguales. Parece que al menos una de las medianas difiere de las demás.

Referencia:

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