LA PRUEBA DE KRUSKAL – WALLIS

Maestría en Estadística – UNSAAC
Diseño de Experimentos

Cuando no se cumple el supuesto de Normalidad de los errores, una alternativa es usar Métodos no paramétricos, como la prueba de Kruskal–Wallis. Para realizar esta prueba se sigue el siguiente procedimiento.

Datos

Los datos consisten de k muestras aleatorias, posiblemente de tamaños diferentes. La i-ésima muestra aleatoria de tamaño n_i se denota por $X_{i1}, X_{i2}, \ldots, X_{in_i}$. Entonces, los datos pueden ser arreglados en columnas como sigue:

Muestra 1	Muestra 2	$\cdots$	Muestra k
X_{11}	X_{21}	$\cdots$	X_{k1}
X_{12}	X_{22}	$\cdots$	X_{k2}
$\vdots$	$\vdots$		$\vdots$
X_{1n_1}	X_{2n_2}	$\cdots$	X_{kn_k}

Sea N que denota el número total de observaciones

$$N = \sum_{i=1}^{k} n_i. \tag{1}$$

Asignar el *rango* 1 al más pequeño de la totalidad de N observaciones, rango 2 al segundo más pequeño, y así sucesivamente hasta el más grande de todas las observaciones, al cual se le asigna rango N. Sea $R(X_{ij})$ que representa el rango asignado a X_{ij}. Sea la *suma de rangos* asignada a la i-ésima muestra

$$R_i = \sum_{j=1}^{n_i} R(X_{ij}), \qquad i = 1, 2, \ldots, k. \tag{2}$$

Calcule R_i para cada muestra. En caso de que existan empates, asignar el promedio de los rangos a cada una de las observaciones empatadas.

Asunciones

1. Todas las muestras son muestras aleatorias de sus respectivas poblaciones.

2. Las muestras son mutuamente independientes.

3. La escala de medida es al menos ordinal.
4. Las k poblaciones tienen funciones de distribución idénticas, o algunas de las poblaciones tiende a producir valores más grandes que otras poblaciones.

Hipótesis

H_0 : Las k poblaciones tienen funciones de distribución idénticas.
H_1 : Al menos una de las k poblaciones tiende a producir observaciones más grandes que las restantes.

Nota: La prueba de Kruskal–Wallis está diseñada para ser más sensible contra la diferencia entre medias en k poblaciones; por esta razón, algunas veces la hipótesis alternativa se expresa como sigue:

$$H_1 : \text{Las } k \text{ poblaciones no tienen medias idénticas.}$$

Estadístico de prueba

El estadístico clásico de Kruskal–Wallis se define como

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1). \tag{3}$$

Cuando existen empates, se aplica la corrección por empates usando

$$C = 1 - \frac{\sum_t (t^3 - t)}{N^3 - N}, \tag{4}$$

donde la suma es sobre todos los grupos de empates de tamaño t. El estadístico corregido es

$$H_c = \frac{H}{C}. \tag{5}$$

Regla de decisión

Si $k = 3$, todos los tamaños muestrales son 5 o menos y no hay empates, el cuantil exacto puede obtenerse de tablas especializadas (p.ej., Tabla A8). Tablas más completas pueden ser obtenidas de Iman, Quade y Alexander (1975). Cuando hay empates o cuando tablas exactas no están disponibles, se usa la aproximación asintótica

$$H_c \overset{H_0}{\sim} \chi^2_{(k-1)}, \tag{6}$$

y se **rechaza** H_0 si

$$H_c > \chi^2_{1-\alpha,\, k-1}. \tag{7}$$

Comparaciones múltiples

Sólo si la hipótesis nula es rechazada, se puede usar el siguiente procedimiento para determinar qué pares de poblaciones tienden a ser diferentes. Para ver si existe diferencia entre las poblaciones i y j a un nivel de significación α, se compara

$$\left|\bar{R}_i - \bar{R}_j\right| \quad \text{con} \quad z_{1-\alpha^*/2}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}, \tag{8}$$

donde $\bar{R}_i = R_i/n_i$ y α^* es un ajuste (por ejemplo Bonferroni) para comparaciones múltiples. Si la diferencia absoluta de medias de rangos excede el umbral, se concluye que las funciones de distribución de las poblaciones i y j difieren al nivel de significación establecido.

Ejemplo No 1

Para comparar tres métodos de enseñanza de programación en cierto lenguaje en computadora:

- **Método A**: instrucción directa con la computadora.
- **Método B**: instrucción de teoría y práctica en computadora.
- **Método C**: sólo clases teóricas.

Se extraen de grandes grupos de alumnos instruidos por uno de los métodos en forma aleatoria: 4 alumnos del método A, 6 del método B y 5 del método C. Luego se les tomó una prueba. Los puntajes obtenidos se muestran a continuación.

A	**B**	**C**
73	91	72
77	90	76
67	81	79
71	83	77
	84	78
	83	

Con los rangos asignados a cada observación:

A		**B**		**C**	
Obs.	**Rango**	**Obs.**	**Rango**	**Obs.**	**Rango**
73	4	91	15	72	3
77	6.5	90	14	76	5
67	1	81	10	79	9
71	2	83	11.5	77	6.5
		84	13	78	8
		83	11.5		
R_i	13.5		75		31.5
n_i	4		6		5

$N = 15$.

Cálculo en R

```
pvalue <- 1 - pchisq(11.11535, 2)
pvalue
# [1] 0.003857735
```

La prueba resultó altamente significativa; se rechaza H_0 de que los tres métodos de enseñanza producen puntajes que se ajustan a la misma distribución.

Como se rechazó H_0 en la prueba anterior, se prosigue con las **comparaciones múltiples** para determinar entre qué métodos existen diferencias en los rendimientos:

Comparación	**Diferencia de medias de rangos**	**Estadístico**
A vs. B	9.125	3.078296*
A vs. C	2.925	3.199059 ns
B vs. C	6.200	2.887698*

Uso de la librería `agricolae`

```
curso <- read.table("calif.txt", T)
calificacion <- curso[,1]
metodo <- curso[,2]

library(agricolae)
kw <- kruskal(calificacion, metodo)
kw
```

```
$statistics
     Chisq     p.chisq
  11.11532 0.003857788

$parameters
  Df ntr  t.value alpha           test name.t
   2   3 2.178813  0.05 Kruskal-Wallis método

$means
    rank calificación      std r Min Max
a  3.375     72.00000 4.163332 4  67  77
b 12.500     85.33333 4.131182 6  81  91
c  6.300     76.40000 2.701851 5  72  79

$comparison
NULL

$groups
  trt  means M
1   b 12.500 a
2   c  6.300 b
3   a  3.375 b
```

```
kw1 <- kruskal(calificacion, metodo, group = FALSE)
kw1
```

```
Study:
Kruskal-Wallis test's
Ties or no Ties

Value: 11.11532
degrees of freedom: 2
Pvalue chisq  : 0.003857788

método,  means of the ranks
  calificación replication
a        3.375           4
b       12.500           6
c        6.300           5

Comparison between treatments mean of the ranks
      Difference   pvalue sig        LCL        UCL
a - b     -9.125 0.000032 *** -12.203296 -6.0467039
a - c     -2.925 0.069606   .  -6.124059  0.2740591
b - c      6.200 0.000534 ***   3.312302  9.0876977
```

Conclusiones

Se ha encontrado diferencias altamente significativas entre las medias de los rendimientos obtenidos con los métodos de enseñanza A y B, así como entre B y C. Por último, sólo se obtuvo diferencia a un nivel de significación del 10 % entre A y C.