Pruebas no paramétricas: Wilcoxon para muestras pareadas y Kruskal-Wallis

F. Aravena[1], M. Elizondo[1], J. Zuñiga[2]

Estadística no paramétrica, Universidad Católica del Maule

Introducción

En investigación estadística, muchas veces los datos no cumplen los supuestos de **normalidad** o **homocedasticidad** requeridos por pruebas paramétricas las **pruebas no paramétricas** son herramientas esenciales para analizar datos *ordinales*, con *outliers* o de *distribuciones desconocidas*.

Este póster presenta dos pruebas fundamentales:

- **Prueba de Wilcoxon para muestras pareadas:**
  - Se usa para comparar dos grupos relacionados (ej.: mediciones antes-después).
  - Es la alternativa no paramétrica al *t-test* para muestras dependientes.
- **Prueba de Kruskal-Wallis:**
  - Diseñada para comparar tres o más grupos independientes.
  - Es la versión no paramétrica del *ANOVA* de una vía.

Objetivo

Proporcionar una guía al aplicar e interpretar estas pruebas en contextos reales, resaltando su utilidad cuando los datos no cumplen supuestos paramétricos.

Prueba de Wilcoxon para pruebas pareadas

La prueba de Wilcoxon para muestras apareadas (también conocida como *Wilcoxon signed-rank test*) propuesta por Frank Wilcoxon en 1945, es una prueba no paramétrica utiliza para comparar dos mediciones relacionadas, como puntajes antes y después de un tratamiento.

A diferencia del t-test pareado, esta prueba no evalúa diferencias de medias, sino de medianas, utilizando los rangos de las diferencias. Es adecuada cuando los datos son ordinales o continuos con outliers, y especialmente útil con muestras pequeñas ($n < 20$). Para tamaños mayores ($n \geq 20$), puede aplicarse una aproximación normal mediante el teorema central del límite. [1]

Supuestos:

- No requiere normalidad
- Muestras pareadas o dependientes.
- Variable continua u ordinal.

Aplicaciones comunes:

- *Estudios antes-después:* Evaluar el efecto de un tratamiento médico.
- *Comparación de dos condiciones en los mismos sujetos:* Por ejemplo, presión arterial en reposo vs. estrés.

Hipótesis

Caso	Hipótesis nula (H_0)	Hipótesis alternativa (H_1)
A)	$\theta_y = \theta_x$	$\theta_y \neq \theta_x$
B)	$\theta_y \leq \theta_x$	$\theta_y > \theta_x$
C)	$\theta_y \geq \theta_x$	$\theta_y < \theta_x$

Cuadro: Formulaciones de hipótesis para prueba de Wilcoxon para muestras pareadas

Procedimiento manual: Prueba de Wilcoxon

1. **Calcular diferencias:** Para cada par de observaciones, se calcula:
$$D_i = X_{i1} - X_{i2}, \quad i = 1, 2, \ldots, n$$
2. **Eliminar diferencias cero:** Si alguna diferencia es $D_i = 0$, se excluye del análisis.
3. **Ordenar valores absolutos:** Se rankean los valores $|D_i|$ de menor a mayor, asignando rangos.
Nota: Si existen dos o más diferencias con igual magnitud y dirección, se asigna a cada una el rango promedio que les correspondería si no hubieran sido iguales. [2]
4. **Asignar signos:** A cada rango se le asocia el signo de D_i.
5. **Calcular estadístico:**
$$T = \sum_{i=1}^{n} R_i^+, \quad \text{donde} \quad R_i^+ = \begin{cases} 0 & \text{si } D_i \leq 0 \\ \text{rango de } |D_i| & \text{si } D_i > 0 \end{cases}$$
6. **Regla de decisión:** Los valores críticos se obtienen a través de la tabla de Wilcoxon considerando n^* (número de diferencias distintas de cero) y α. La regla de decisión es:
   - **a)** Rechazar H_0 si $T < T_{\alpha/2}(n^*)$ o $T > T_{1-\alpha/2}(n^*)$
   - **b)** Rechazar H_0 si $T > T_{1-\alpha}(n^*)$
   - **c)** Rechazar H_0 si $T < T_{\alpha}(n^*)$

Ejercicio de la prueba de Wilcoxon

Un grupo de psicólogos evaluó a personas próximas a jubilarse, con el objetivo de determinar si su salud mental tiende a ser inferior después del retiro. Para ello, se midió el puntaje de salud mental antes y después del retiro en una muestra de individuos.

$$H_0 : \theta_Y \geq \theta_X \quad H_a : \theta_Y < \theta_X$$

Individuo	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Después (Y)	70	75	84	90	81	95	87	72	92	85	88	76	85	81	84
Antes (X)	76	80	86	87	85	95	97	75	87	96	98	77	80	87	89
$Y - X$	-6	-5	-2	+3	-4	0	-10	-3	+5	-11	-10	-1	+5	-6	-5
Rango	10.5	7.5	3.5	5	7.5		12.5	3.5	10.5	14	12.5	1	10.5	10.5	7.5
Signo	-	-	-	+	-		-	-	+	-	-	-	+	-	-

Cuadro: Prueba de Wilcoxon. *Fuente:* [1].

Consecuentemente, tenemos que:
$$T = 18{,}5 \quad \text{y} \quad n^* = 14$$

Siendo $\alpha = 0{,}05$, el valor crítico de Wilcoxon es:
$$W_{0{,}05}(14) = 26$$

Como $T = 18{,}5 < 26$, se **rechaza la hipótesis nula** H_0.
Se implemento R utilizando la función `wilcox.test()`.

Estadístico V	p-valor
18.5	0.0349

Cuadro: Resultados de la prueba de Wilcoxon en R

Conclusión: Con un nivel de significancia del 5 %, se concluye que el nivel de salud mental de la población tiende a ser **inferior después del retiro**.

Prueba de Kruskal-Wallis

La prueba de Kruskal-Wallis es una técnica no paramétrica que se utiliza para comparar tres o más grupos independientes cuando la variable de interés es ordinal o continua y los datos no cumplen con el supuesto de normalidad. Evalúa si existen diferencias significativas entre las medianas de los grupos, basándose en rangos en lugar de los valores originales.

Es ideal para muestras pequeñas, datos sesgados o con valores atípicos.

Supuestos

- No requiere normalidad
- Los grupos deben ser independientes
- Se aplica a tres o más grupos
- Variable dependiente ordinal o continua

Hipótesis

Hipótesis nula (H_0)	No hay diferencias significativas entre los grupos:
	$\mu_1 = \mu_2 = \cdots = \mu_k$
Hipótesis alternativa (H_1)	Al menos una mediana es diferente:
	$\mu_i \neq \mu_j$ para algún $i \neq j$

Cálculo del estadístico:

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \left(\frac{R_j^2}{n_j} \right) - 3(N+1)$$

- N: número total de observaciones en todos los grupos
- n_j: número de observaciones en el grupo j
- k: número de grupos
- R_j: suma de los rangos en el grupo j

Región de rechazo:

- Se compara H con el valor crítico de una χ^2 con $k - 1$ grados de libertad:
  - Si $H > \chi^2_{k-1}$: se rechaza H_0
  - Si $H \leq \chi^2_{k-1}$: no se rechaza H_0
- También se puede usar el valor-p:
  - Si p-valor $< \alpha$: se rechaza H_0
  - Si p-valor $\geq \alpha$: no se rechaza H_0

Ejercicio de la prueba de Kruskal-Wallis

Un investigador desea comparar la efectividad de **cuatro tratamientos distintos** (A, B, C y D) aplicados a un grupo de pacientes.

Datos observados:

Tratamiento A	R_A	Tratamiento B	R_B	Tratamiento C	R_C	Tratamiento D	R_D
68	**1**	80	**13**	74	**9**	92	**20**
72	**6.5**	85	**15**	70	**3.5**	88	**17**
75	**10**	78	**11**	69	**2**	90	**19**
70	**3.5**	83	**14**	72	**6.5**	89	**18**
73	**8**	79	**12**	71	**5**	87	**16**
R_j	**29**		**65**		**26**		**90**

Cuadro: Valores observados, rangos asignados y suma de rangos por grupo (A, B, C, D).

Prueba de Kruskal-Wallis: Cálculo e interpretación

Hipótesis:

$$H_0 : \mu_A = \mu_B = \mu_C = \mu_D \qquad H_a : \mu_i \neq \mu_j \quad \text{para algún par } i \neq j$$

Estadístico de prueba:

$$H = \frac{12}{20(20+1)} \left(\frac{29^2 + 65^2 + 26^2 + 90^2}{5} \right) - 3(20+1)$$
$$H \approx 16{,}18$$

Valor-p:

$$p = 1 - P\left(\chi^2_{(3)} \leq 16{,}18\right) \approx 0{,}001$$

Se implemento R utilizando la función `kruskal.test()`.

Estadístico H	Grados de libertad (gl)	p-valor
16.121	3	0.0011

Cuadro: Resultados de la prueba de Kruskal-Wallis

Conclusión: Con un nivel de significancia del 5 %, se **rechaza** H_0. Existe suficiente evidencia muestral para concluir que al menos uno de los tratamientos difiere significativamente de los demás.

Conclusiones

- **Wilcoxon** es óptima para diseños antes-después o experimentales con muestras dependientes.
- **Kruskal-Wallis** es la opción preferida para comparar múltiples tratamientos, categorías o poblaciones independientes.
- Ambas pruebas son fáciles de implementar en **RStudio**, utilizando funciones integradas como `wilcox.test()` y `kruskal.test()`, junto con visualizaciones claras como *boxplots* o gráficos de perfiles.

Referencias

[1] Martha Elva and Quito López Tirado.
Métodos estadísticos no paramétricos.
Universidad Autónoma Chapingo, México, 1 edition, 1993.

[2] Mario Miguel Ojeda, Alberto Castillo, and Julián Felipe Díaz.
Principios de estadística no paramétrica: metodología con ilustraciones.
Editorial Académica Española, España, 2013.