

Prueba de hipótesis en R: Prueba t de Student y prueba de Wilcoxon

Hypothesis testing in R: Student's t test and Wilcoxon test

Nolbert González Hernández¹ (nolbertreblon@gmail.com) (<https://orcid.org/0000-0002-9579-1073>)

Angel Reyes González² (angelreyesgonzalez46@gmail.com) (<https://orcid.org/0000-0001-9844-8839>)

Resumen

Las pruebas estadísticas son fundamentales en la investigación científica porque proporcionan un marco riguroso y objetivo para analizar y comparar datos. Dentro de estas pruebas estadísticas, la Prueba t de Student y la prueba de Wilcoxon son comúnmente utilizadas. Sin embargo, entre estas dos pruebas existen diferencias significativas, principalmente relacionadas con los tipos de datos en los que se aplican. La elección entre la prueba t de Student y la prueba de Wilcoxon puede constituir un problema en algunas ocasiones, cuya solución depende de las características de los datos. En esta dirección, el objetivo de la presente investigación es demostrar cómo se aplican la prueba t de Student y la prueba de Wilcoxon empleando el software R, mediante la ilustración de ejemplos prácticos. Para ello se emplearon varios métodos tales como el análisis-síntesis, inducción-deducción y la sistematización teórica, los cuales permitieron realizar un estudio procedimental de la prueba t de Student y la prueba de Wilcoxon. A partir de lo enunciado, la investigación ofrece una guía práctica para aplicar la prueba t de Student y la prueba de Wilcoxon utilizando el software R. En ese sentido, se proporciona una explicación detallada de cómo realizar ambas pruebas en R, incluyendo las funciones y las interpretaciones de los datos. El principal resultado de la investigación consiste en ejemplos prácticos y concretos que ayudan a los lectores a comprender mejor cómo aplicar estas pruebas estadísticas en los procesos de investigación científica.

Palabras clave: prueba t de student, prueba de Wilcoxon, análisis estadístico y gráfico.

Abstract

Statistical tests are essential in scientific research because they provide a rigorous and objective framework for analyzing and comparing data. Within these statistical tests, the Student t-test and the Wilcoxon test are commonly used. However, between these two tests, there are significant differences, mainly related to the types of data to which they are applied. The choice between the Student t-test and the Wilcoxon test can sometimes be a problem, the solution to which depends on the characteristics of the data. In this direction, the objective of this research is to demonstrate how the Student t-

¹ Máster en Educación Matemática Universitaria. Universidad de Holguín, Cuba.

² Especialista en Primer Grado de Medicina Interna. Universidad de Ciencias Médicas de Holguín, Cuba.

test and the Wilcoxon test are applied using the R software, by illustrating practical examples. For this, several methods were used such as analysis-synthesis, induction-deduction, and theoretical systematization, which allowed a procedural study of the Student's t-test and the Wilcoxon test to be carried out. From the above, the research offers a practical guide to applying the Student t-test and the Wilcoxon test using the R software. In this sense, a detailed explanation of how to perform both tests in R is provided, including the functions and interpretations of the data. The main result of the research consists of practical and concrete examples that help readers better understand how to apply these statistical tests in scientific research processes.

Key words: student t test, Wilcoxon test, statistical analysis and graph

Introducción

Las pruebas estadísticas son fundamentales en la investigación científica, al constituirse en una vía que permite analizar y comparar datos de forma objetiva y rigurosa (Cardoso et al., 2022 y De la Lama et al., 2022). Entre las pruebas comúnmente empleadas se encuentran: la Prueba t de Student y el test de Wilcoxon, las cuales permiten comparar medias y determinar si existen diferencias significativas entre dos conjuntos de datos (López et al., 2020 y Machado et al., 2019).

Para facilitar estos procesos de análisis estadísticos se han concebido varias herramientas informáticas que lo facilitan que se apoyan en el empleo del software R, entre ellas: la Prueba t de Student y prueba de Wilcoxon.

El software R se reconoce como un lenguaje de programación de código abierto, con accesibilidad y compatibilidad con múltiples sistemas operativos y múltiples bibliotecas, los cuales permiten a los usuarios personalizar y extender su funcionalidad para satisfacer necesidades específicas (Rode y Ringel, 2019).

La Prueba t de Student, por su parte, también es considerada como una prueba paramétrica que se utiliza para determinar si hay una diferencia significativa entre dos muestras. Además, esta prueba estadística puede ser empleada en disímiles campos de la investigación científica (De Winter, 2019 y Prabhaker et al., 2019). En esta prueba, la hipótesis nula es una afirmación de que los efectos que se observan en las muestras se deben a una casualidad, es decir, no existen diferencias entre las muestras o no hay efecto de la variable independiente sobre la dependiente.

Por las razones antes expuestas, se plantea como objetivo de esta investigación demostrar cómo se aplican la prueba t de Student y la prueba de Wilcoxon empleado el software R mediante la ilustración de ejemplos prácticos.

Materiales y métodos

Para el desarrollo de la investigación se utilizaron como métodos teóricos el análisis-síntesis, inducción-deducción, y la sistematización teórica que permitieron realizar un estudio procedimental de ambos métodos estadísticos. En esta dirección, estos

métodos fueron empleados en la revisión de la literatura existente sobre el uso de la prueba t de Student y de la prueba de Wilcoxon, sus limitaciones y aplicaciones. Además, resultaron de utilidad en la organización, clasificación y la interpretación de los resultados.

Con particular énfasis, se empleó el método estadístico-matemático para realizar un análisis de los resultados alcanzados con la implementación de ambas pruebas estadísticas. Los datos utilizados en este estudio fueron generados de forma tal que cumplan los supuestos requeridos para la aplicación de cada prueba de hipótesis. Una vez generados los datos, se procedió a aplicar las pruebas.

Resultados

En el caso del Software R a continuación se describe como hacer uso del mismo para la obtención de una estadística objetiva y confiable a partir de las herramientas seleccionadas.

Para introducir las muestras en R se puede definir dos variables (grupo 1 y grupo 2) donde cada variable es un vector que contiene los datos obtenidos, separados por una coma:

Definir variables

```
grupo1
```

```
= c(82.19,83.84,92.79,85.35,85.64,93.57,87.30,78.67,81.56, 82.77,91.12,86.79,87.00,85.55,82.22,93.93,
```

```
87.48,75.16,88.50,82.63,79.66,83.91,79.86,81.35,81.87,76.56,89.18,85.76, 79.30,91.26)
```

```
grupo2
```

```
= c(82.13,78.52,84.47,84.39,84.10,83.44,82.76,79.69,78.47,78.09,76.52,78.96,73.67,90.84, 86.03,74.38,
```

```
77.98,77.66,83.89,79.58,81.26,79.85,79.78,86.84,78.87,87.58,72.25,82.92,80.61,81.07)
```

Es importante considerar que para aplicar una prueba t los datos deben cumplir con las siguientes suposiciones:

- Independencia: las observaciones deben ser independientes entre sí. Esto significa que la presencia de una observación no afecta la presencia de otra.
- Normalidad: la distribución de los datos debe ser normal.

Para comprobar la normalidad de los datos en R se puede utilizar un gráfico Q-Q (*quantile-quantile*) o una prueba de *Shapiro-Wilk*, entre otros (Flores y Flores, 2021). El gráfico $Q - Q$ compara cuantiles de los datos con las de una distribución normal. Por lo que, si los datos son normalmente distribuidos los puntos en el gráfico $Q - Q$ deben

formar una línea recta. En R, se puede utilizar la función `qqnorm()` para observar esta distribución de los datos (figura 2):

```
qqnorm(grupo1)
```

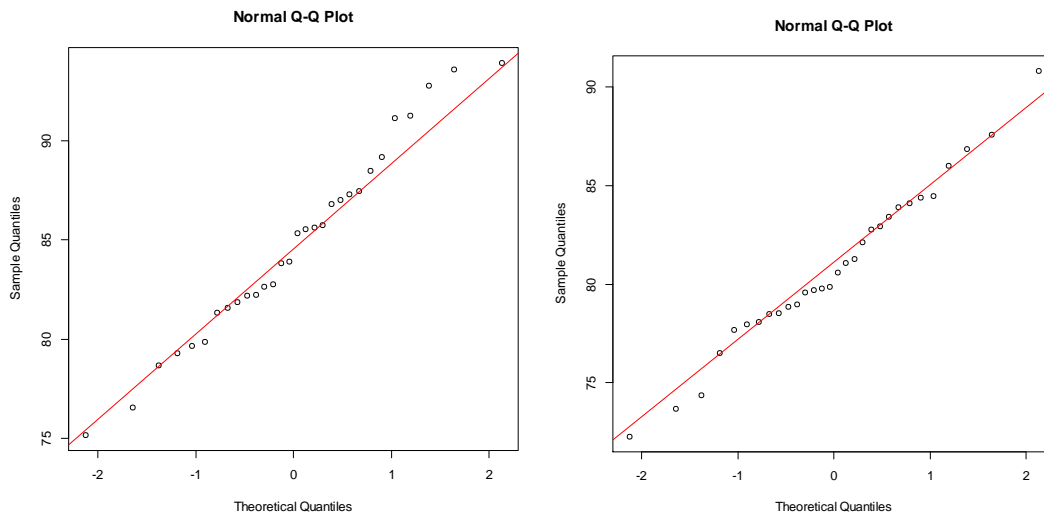
```
qqline(grupo1,col = "red")
```

```
qqnorm(grupo2)
```

```
qqline(grupo2,col = "red")
```

Figura 1

Gráfico $Q - Q$ de la muestra 1 y gráfico $Q - Q$ de la muestra 2



Fuente: elaboración propia utilizando R

En el gráfico $Q - Q$ de las dos muestras, algunos datos se “salen” ligeramente de la línea. Esto puede indicar que los datos no son completamente normales. Sin embargo, la normalidad es una propiedad difícil de medir con una alta precisión. Por lo que, aunque los datos no son perfectamente normales se puede utilizar esta prueba siempre y cuando las diferencias en la distribución no sean demasiado grandes. Esto se debe principalmente a que la prueba t es robusta (capacidad para producir resultados válidos y confiables incluso cuando se incumplen algunos de los supuestos básicos que se requieren para su aplicación) y pueden funcionar bien, incluso cuando los datos no son perfectamente normales.

La prueba de *Shapiro-Wilk* es una prueba estadística que se puede usar para probar la normalidad de una muestra. Si el valor p de la prueba es menor que **0.05**, entonces la distribución en cuestión se considera significativamente diferente de una distribución normal. En R se puede aplicar esta prueba utilizando la función `shapiro.test()` como se muestra a continuación (figura 3).

```
sha <- shapiro.test(grupo1)
```

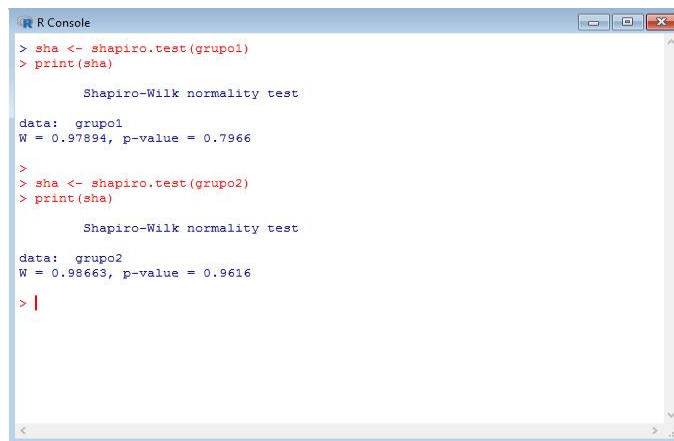
```
print(sha)
```

```
sha <- shapiro.test(grupo2)
```

```
print(sha)
```

Figura 2.

Resultados de la prueba Shapiro-Wilk exportada de la consola de R



```
R Console
> sha <- shapiro.test(grupo1)
> print(sha)
      Shapiro-Wilk normality test
data:  grupo1
W = 0.97894, p-value = 0.7966
>
> sha <- shapiro.test(grupo2)
> print(sha)
      Shapiro-Wilk normality test
data:  grupo2
W = 0.98663, p-value = 0.9616
> |
```

Fuente: elaboración propia utilizando R.

Como $p - value$ es mayor que **0.05** para ambos grupos, se puede asumir que los datos siguen una distribución normal.

Varianza igual: se asume que las varianzas de las poblaciones de las que provienen las dos muestras son iguales. La varianza es una medida de cuánto varían los datos alrededor de la media, es una forma de entender cuánto de dispersos están los datos.

La función `var()` en R permite obtener la varianza (figura 4).

```
varianza1 <- var(grupo1)
```

```
print(varianza1)
```

```
varianza2 <- var(grupo2)
```

```
print(varianza2)
```

Es importante tener en cuenta que la varianza es una medida en cuadrados, por lo que si se quiere entender la dispersión en las mismas unidades que los datos, se necesita tomar la raíz cuadrada de la varianza para obtener la desviación estándar. En estos casos particulares, la desviación estándar sería 4.12 y de 4.9 respectivamente.

Figura 3.

Resultados de la prueba de varianza exportados de la consola de R



```
R Console
> varianza1 <- var(grupo1)
> print(varianza1)
[1] 24.06231
> varianza2 <- var(grupo2)
> print(varianza2)
[1] 17.43451
> |
```

Fuente: elaboración propia utilizando R.

Los resultados obtenidos significan que, en promedio, cada punto de datos en el primer grupo está a 4.12 unidades de la media, y cada punto de datos en el segundo grupo está a 4.9 unidades de la media. Teniendo en cuenta los resultados obtenidos respecto a la normalidad y a la varianza, se considera si es viable aplicar o no la prueba t de Student.

La función `t.test()` en R realiza la prueba t de Student y devuelve los resultados de la prueba. Si el valor *p* es menor que el nivel de significancia que se ha definido (5% por defecto) entonces, se puede rechazar la hipótesis nula y concluir que las diferencias que se observan entre los grupos son estadísticamente significativas. Es decir, es probable que no sea resultado de la variabilidad aleatoria en los datos.

Para realizar la prueba en R se utiliza la siguiente función:

```
# Realizar la prueba t de Student
```

```
resultado <- t.test(grupo1, grupo2)
```

```
# Imprimir el resultado
```

resultado

Figura 4.

Resultados exportados de la consola de R

```

R Console
> # Definir variables
> grupo1= c(82.19,83.84,92.79,85.35,85.64,93.57,87.30,78.67,81.56,82.77,91.12,85
+ 87.48,75.16,88.50,82.63,79.66,83.91,79.86,81.35,81.87,76.56,89.18,85.76,79.30$
> grupo2=c(82.13,78.52,84.47,84.39,84.10,83.44,82.76,79.69,78.47,78.09,76.52,78$
+ 77.98,77.66,83.89,79.58,81.26,79.85,79.78,86.84,78.87,87.58,72.25,82.92,80.61$
> # Realizar la prueba t de Student
> resultado <- t.test(grupo1,grupo2)
> # Imprimir el resultado
> resultado

Welch Two Sample t-test

data: grupo1 and grupo2
t = 3.2925, df = 56.557, p-value = 0.001715
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.516820 6.227846
sample estimates:
mean of x mean of y
84.75900 80.88667

> |
  
```

Fuente: elaboración propia utilizando R.

Un valor p de **0.03** significa que existe solo un 3% de probabilidad de que se obtengan estos resultados si la hipótesis nula fuera cierta. En este ejemplo, como **0.001715** es menor que **0.05** se puede rechazar la hipótesis nula y concluir que las medias de los dos grupos son significativamente diferentes.

En el caso del Test de Wilcoxon se emplea como una prueba estadística no paramétrica que se utiliza para comparar dos grupos emparejados. Este test se puede realizar utilizando dos técnicas diferentes, el test de rango con signo y el test de suma de rangos (Ramírez y Polack, 2020). El test de rango con signo se utiliza cuando los datos emparejados son de naturaleza ordinal, mientras que el test de suma de rangos de Wilcoxon se utiliza cuando los datos emparejados son de naturaleza continua.

En R, se puede realizar este test empleando la función `wilcox.test()` para comparar dos muestras emparejadas. En el caso de muestras apareadas, en los datos se deben especificar los dos vectores correspondientes (figura 5).

Introducir vectores con los datos

antes = c(2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4)

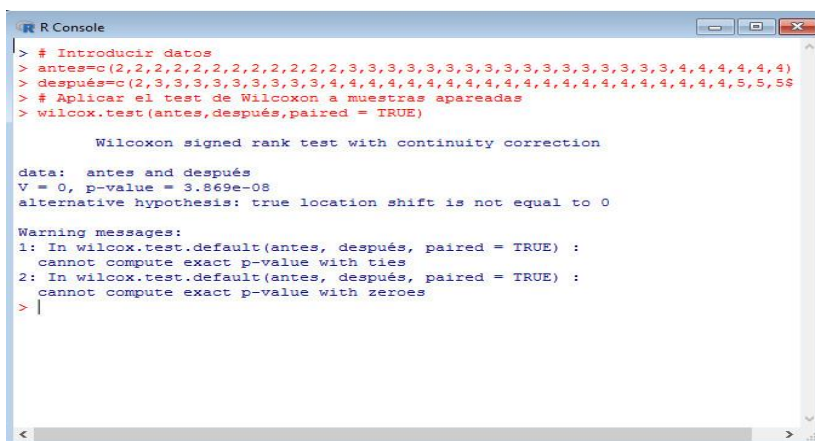
después = c(2,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,5,5,5,5)

Aplicar el test de Wilcoxon a muestras apareadas

```
wilcox.test(antes, después, paired = TRUE)
```

Figura 5.

Resultados de una prueba de Wilcoxon exportados de la consola de R



```
R Console
> # Introducir datos
> antes=c(2,2,2,2,2,2,2,2,2,2,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4)
> después=c(2,3,3,3,3,3,3,3,3,3,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,4,5,5,5,5)
> # Aplicar el test de Wilcoxon a muestras apareadas
> wilcox.test(antes,después,paired = TRUE)

      Wilcoxon signed rank test with continuity correction

data:  antes and después
V = 0, p-value = 3.869e-08
alternative hypothesis: true location shift is not equal to 0

Warning messages:
1: In wilcox.test.default(antes, después, paired = TRUE) :
   cannot compute exact p-value with ties
2: In wilcox.test.default(antes, después, paired = TRUE) :
   cannot compute exact p-value with zeroes
> |
```

Fuente: elaboración propia utilizando R.

El valor de *p* indica que la probabilidad de obtener un resultado similar al observado si la hipótesis nula es verdadera, es extremadamente baja. Si el valor de *p* es menor que el nivel de significación predefinido (por defecto 0.05), se rechaza la hipótesis nula y se acepta la hipótesis alternativa. En este caso, el valor de *p* es menor que **0.05**, lo que sugiere que hay evidencia suficiente para rechazar la hipótesis nula y aceptar la hipótesis alternativa.

Discusión de los resultados

Comparación entre ambas pruebas

La Prueba *t* de Student como prueba paramétrica se utilizó cuando se cumplieron los supuestos de *independencia*, *normalidad* y *varianza igual* en los datos, mientras que la Prueba de Wilcoxon no es una prueba paramétrica y se utilizó cuando no se cumplieron estos supuestos. En ambas pruebas, la hipótesis nula (*H*0) ofreció una afirmación que no se apreció diferencia y relación entre las variables. Mientras que la hipótesis alternativa (*H*1) establece lo contrario.

Se utilizaron las funciones `t.test()` y `wilcox.test()` para realizar una prueba *t* de *student* o una Prueba de *Wilcoxon* respectivamente. Para interpretar los resultados de ambas pruebas, se comparó el valor *p* (*p – value*) respecto al nivel de significancia deseado (predefinido como **0.05**). Por lo que, si el valor *p* es menor que el nivel de significancia, se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

Es importante destacar que la interpretación de los resultados de cada prueba de hipótesis se realizó en un contexto específico. En consecuencia, aun cuando los resultados no garantizan la verdad absoluta, sí proporcionan evidencia para aceptar o refutar hipótesis en las investigaciones.

En los datos en los que se cumplieron los supuestos de normalidad y homogeneidad de varianzas, se procedió a aplicar la prueba *t* de Student para comparar las medias de los datos bajo estudio. Los resultados obtenidos proporcionaron evidencia para respaldar la hipótesis alternativa, la cual afirmaba que existía diferencia entre los dos grupos.

En los datos no paramétricos se recurrió a la prueba de Wilcoxon para comparar las medianas de los grupos. Los resultados proporcionaron evidencia para respaldar la hipótesis alternativa, la cual significa que existe una diferencia entre los dos grupos.

Conclusiones

En la investigación se describe y aplica de manera rigurosa las pruebas de hipótesis *t* de Student y la prueba de Wilcoxon utilizando el software R. El principal aporte de la investigación consiste en realizar un estudio procedimental de estas pruebas. La investigación contribuye a la comprensión de las fortalezas y limitaciones de las pruebas de hipótesis aplicadas, lo que permite a los investigadores seleccionar la correcta según las características de sus datos y los objetivos de su estudio.

Se promueve la estandarización en la aplicación de estas pruebas de hipótesis utilizando funciones disponibles en R. Además, al familiarizar a los lectores con el uso de funciones en R, se les capacita para explorar técnicas similares en futuras investigaciones.

Referencias bibliográficas

Cardoso, L., Castro, G., Fernández, C. (2022). La Estadística en función de la investigación educativa. Reto para los profesionales de la educación. *Mendive*, 20(1), 270-284. <https://dialnet.unirioja.es/servlet/articulo?codigo=8317913>

De la Lama, P., de la Lama, M., de la Lama, A. (2022). Los instrumentos de la investigación científica. *Horizonte de la Ciencia*, 12(22). <https://www.redalyc.org/journal/5709/570969250014/570969250014.pdf>

- De Winter, J. (2019). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research, and Evaluation*, 18. <https://doi.org/10.7275/e4r6-dj05>
- Flores, C. y Flores, K. (2021). Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. *Societas*, 23(2), 83–106. <https://matriculapre.up.ac.pa/index.php/societas/article/view/2302>
- López, R., Cabrera, E. y Palmero, D. (2020). Pertinencia de métodos estadísticos empleados para medir el clima organizacional en el ámbito pedagógico. *MediSur*, 18(5), 780-788. http://scielo.sld.cu/scielo.php?pid=S1727-897X2020000500780&script=sci_arttext
- Machado, W. y Rodríguez, M. (2019). Criterios para la selección y aplicación de herramientas estadísticas en la investigación educativa. *Delectus*, 2(1). <http://portal.amelica.org/ameli/journal/390/3902682005/html/>
- Prabhaker, M., Uttam, C., Priyadarshni, M., and Gaurav, P. (2019). Application of Student's t-test. *Analysis of Variance and Covariance*, 22(4), 407-411. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6813708/>
- Ramírez, A. y Polack, A. (2020). Estadística inferencial. Elección de una prueba estadística no paramétrica en investigación científica. *Horizonte de la Ciencia*, 10(19), 191–208. <https://doi.org/10.26490/uncp.horizonteciencia.2020.19.597>
- Rode, J. B., y Ringel, M. M. (2019). Statistical Software Output in the Classroom: A Comparison of R and SPSS. *Teaching of Psychology*, 46(4), 319-327. <https://doi.org/10.1177/0098628319872605>

Conflicto de intereses: Los autores declaran no tener conflictos de intereses.

Contribución de los autores: Los autores participaron en la búsqueda y análisis de la información para el artículo, así como en su diseño y redacción.