

Estudio comparativo de potencia estadística en diferentes escenarios de pruebas para la detección de diferencias de medias con múltiples tratamientos

Daniel Alfaro¹; Luis Solano¹; Luis González¹

daniel.alfarofigueroa@ucr.ac.cr, luis.solanocoto@ucr.ac.cr, luiscarlos.gonzalez@ucr.ac.cr

RESUMEN

Este análisis se centró en observar, para distintas distribuciones, la probabilidad de cometer error tipo I con cinco métodos propuestos: la prueba ANOVA, la prueba Kruskal-Wallis y 3 métodos de permutaciones con simulaciones. Para ello, se analizó la potencia de la prueba en diferentes métodos de análisis y con diferentes escenarios para cada distribución. También se quería determinar si el añadir una mayor cantidad de medias diferentes entre tratamientos, sin alterar la detectabilidad de la prueba, aumentaba o disminuía la potencia, y definir, para cada caso de distribución analizada, cuál era el método que obtenía una potencia de la prueba mayor. Se encontró que, ante violaciones a los supuestos de normalidad y homocedasticidad, se altera la significancia real de las pruebas llevadas a cabo con el método de análisis de varianza paramétrico. Además, se determinó que para casos donde se viola el supuesto de normalidad, pero en el que la distribución de los datos no es muy asimétrica ni de colas muy largas, la violación al supuesto no es tan grave al hacer un análisis de varianza. Además, se concluyó que el añadir más medias diferentes entre tratamientos manteniendo constante la detectabilidad de la prueba genera, por lo general, potencias más bajas.

ABSTRACT

This paper focused on observing, in different distributions, the probability of error type I using five analysis methods; ANOVA, Kruskal-Wallis test and three permutation tests. To do so, the statistical power of a test was calculated for different scenarios. Another goal of this investigation was to determine whether or not increasing the amount of means that are different, in the simulated distributions, would affect the statistical power, and define for each scenario which method brings the highest statistical power. Findings suggest that in situations when the normality and homogenous variances assumptions are not met, the ANOVA real test significance is altered. In these same cases, when assumptions are not met, but the distributions are not too asymmetric, or long tailed, this violation of the assumptions does not imply a severe impact in the ANOVA's real test significance. The analysis showed that the more different means between treatments were added, keeping the same detectability level, the tests generate lower statistical power.

PALABRAS CLAVE: potencia, pruebas de hipótesis, supuestos, métodos paramétricos y no paramétricos, error tipo I, prueba de permutaciones.

¹ Estudiante de Bachillerato en estadística de la Universidad de Costa Rica

INTRODUCCIÓN

En los diseños experimentales uno de los objetivos es determinar si los promedios de la variable respuesta difieren al ser agrupados de acuerdo con uno o más factores. Para compararlos se llevan a cabo pruebas de hipótesis que concluyen sobre la igualdad o no de los promedios. Al realizar estas pruebas se definen las hipótesis de la prueba, en estos casos la hipótesis nula indica que los promedios de los tratamientos son todos iguales mientras que la hipótesis alternativa señalaría que al menos uno de los promedios de los tratamientos es diferente a los demás. Estas pruebas se dividen en paramétricas y no paramétricas, estando las paramétricas sujetas a ciertos supuestos, una limitante que no presentan las alternativas no paramétricas. Las conclusiones de estas pruebas se realizan en base a la hipótesis nula, en otras palabras, se busca rechazar o no rechazar la hipótesis nula.

Al tomar decisiones, basándose en la evidencia y la prueba estadística, existe la posibilidad de cometer errores, estos posibles errores son clasificados como errores de tipo I y tipo II, denotados como (α) y (β) respectivamente. La probabilidad de rechazar la hipótesis nula cuando no se debe rechazar se conoce como el error tipo I, se acostumbra fijar una probabilidad máxima de cometer este error. En contraste, se encuentra presente el error tipo II, el cual consiste en el no rechazo de la hipótesis nula cuando en realidad debería ser rechazada. La potencia es una probabilidad opuesta a la probabilidad de cometer el error tipo II y expresa que tanta seguridad tiene el investigador de llegar a la conclusión correcta de rechazar la hipótesis nula cuando las muestras provienen de poblaciones cuyos promedios distan entre sí. Para calcular la potencia de la prueba se define una magnitud de corte a partir de la cual la diferencia empieza a ser “relevante”, esta se llama detectabilidad de la prueba y se denota con δ (Alvarado, 2020). Esta investigación se centró en comparar la potencia estadística y la probabilidad de cometer error tipo I, entre distintos escenarios de distribución de las respuestas y cantidad de diferencias entre las medias, para alternativas paramétricas y no paramétricas de pruebas de igualdad de medias.

Existen diferentes pruebas que son comúnmente utilizadas para determinar si los promedios de la variable respuesta son diferentes al agruparla de acuerdo con los factores. Esta investigación se centró en tres de ellas: el análisis de varianza (ANOVA, siglas en inglés), la prueba Kurskal-Wallis y la prueba de permutaciones con tres estadísticos diferentes. Como alternativa paramétrica se recurrió al análisis de varianza. La prueba F, de un análisis de varianza, es una prueba estadística basada en la distribución de probabilidad F, que se utiliza para comparar las varianzas de 2 o más grupos. El estadístico de la prueba F se obtiene al dividir la varianza “entre grupos” entre la varianza “dentro de los grupos”, si el estadístico F es grande esto sugeriría que los promedios de los diferentes grupos están separados, o que los promedios de los grupos son diferentes (Larson, 2008 & Sawyer, 2009).

De acuerdo con Larson (2008) al modelar en un análisis de varianza se espera que los datos cumplan una serie de supuestos. Estos son el supuesto de independencia de las observaciones, homocedasticidad y por último, el supuesto de análisis para esta investigación, que la distribución de las observaciones de cada grupo de tratamiento sea una distribución normal.

En muchos casos los supuestos mencionados previamente no se cumplen, por lo que hay que recurrir a alternativas no paramétricas. Una de las pruebas no paramétricas más conocidas es la prueba Kruskal-Wallis. Esta prueba se utiliza cuando los datos obtenidos no cumplen con los supuestos de normalidad y homocedasticidad. Se utiliza para determinar si hay diferencias significativas entre las medianas de 3 o más grupos, se calcula a partir de los rangos de los datos en cada grupo y se basa en la suma de los rangos al cuadrado ponderados por el tamaño de cada grupo (Ostertagova, Ostertag, y Kováč, 2014).

Además del ANOVA y la prueba Kruskal-Wallis, en esta investigación se trabajó con la prueba de permutaciones. Esta se basa en la hipótesis nula de que todas las observaciones proceden de la misma población y tiene como objetivo determinar si la diferencia observada entre los grupos se debe a un factor determinado o solo a la variabilidad esperada por la naturaleza de un proceso aleatorio (Alvarado, 2020).

En esta investigación el análisis se desarrolló en torno a tres objetivos específicos. El primero, observar, para distintas distribuciones, la probabilidad de cometer error tipo I con cada uno de los métodos propuestos, con el fin de determinar si se estaba alterando la verdadera significancia de las pruebas. El segundo fue concluir si el añadir medias diferentes entre tratamientos, sin alterar la detectabilidad de la prueba, aumentaba o disminuía la potencia. Y por último, definir, para cada caso de distribución analizada, cuál era el método que obtenía una potencia de la prueba mayor. Fue de interés también analizar si la prueba de Kruskal-Wallis generaba potencias similares a la prueba de permutaciones con el estadístico de Kruskal-Wallis y si el ANOVA generaba potencias similares a prueba de permutaciones con el estadístico F.

METODOLOGÍA

Para el análisis, se propusieron cuatro casos diferentes: uno para una distribución normal, otro para una distribución exponencial, otro para una escala con valores discretos entre 0 y 4, y un último para una escala con valores discretos entre 0 y 9. La parametrización de cada distribución puede ser consultada en los anexos de este documento (ver Tablas 6 a la 10). Los experimentos simulados fueron compuestos por un solo factor con cinco niveles. Se decidió incorporar la distribución normal al análisis para tomar un caso que cumpliera a cabalidad con los supuestos requeridos por el análisis de varianza; además, se simularon las muestras normales con varianza constante entre tratamientos, cumpliendo con el supuesto de homocedasticidad.

La distribución exponencial fue añadida al análisis para representar un caso en el que fuera evidente la violación al supuesto de normalidad y de homocedasticidad (este último para las etapas en las que al menos una de las medias era diferente). Por otro lado, se decidió trabajar con dos escalas diferentes con el objetivo de analizar si habían diferencias en el comportamiento de la potencia de la prueba paramétrica (ANOVA) y de las no paramétricas al usar escalas con dominio mayor. La escala 0-4 fue parametrizada para ser simétrica, mientras que la escala 0-9 fue parametrizada de modo que fuera asimétrica (ver Tablas 9 y 10, y Figuras 1 y 2).

El incorporar la distribución de una escala simétrica permitió tomar en cuenta a distribuciones que, a pesar de no ser normales, son apropiadas para llevar a cabo un análisis de varianza paramétrico, pues se sabe que la prueba F del análisis de varianza es robusta a la violación leve de la normalidad cuando la distribución es simétrica. Por el contrario, llevar a cabo el análisis para la distribución de una escala asimétrica permite tomar en cuenta a un caso en el que el análisis de varianza paramétrico no sea apropiado por la violación del supuesto de normalidad y la asimetría de la distribución.

La decisión de usar una alternativa paramétrica para un conjunto de datos que no cumple con sus supuestos tiene como resultado el obtener inferencias inválidas en los intervalos de confianza y las pruebas de hipótesis desarrolladas, pues se altera la verdadera significancia y verdadera potencia de las pruebas (Alvarado, 2020). Además de la invalidación de las inferencias, fue de interés determinar si existía una afectación en términos prácticos en la potencia de las pruebas al optar por métodos paramétricos cuando no se cumplen sus supuestos. Bajo el mismo entendido, fue de interés probar otros métodos, como pruebas de permutaciones, para determinar si eran más potentes y si

controlaban mejor el error tipo I en muestras obtenidas de poblaciones con distintas parametrizaciones. El proceso seguido por las pruebas de permutaciones es el siguiente: se reasignan aleatoriamente las unidades experimentales a cada uno de los grupos manteniendo el tamaño original de cada uno, seguidamente se calcula el estadístico de preferencia y se almacena el valor, se repiten los pasos anteriores un alto número de veces y por último se calcula el valor p como la proporción de permutaciones en las que, el estadístico de prueba, es igual o mayor al observado en los datos originales (Bautista & Gómez, 2017).

El proceso de simulación fue desarrollado en cinco etapas, cada una compuesta por mil iteraciones. Las etapas fueron definidas por el número de medias reales diferentes entre los tratamientos. Es decir, para cada distribución, en la primera etapa se generaron muestras con todas las medias iguales entre tratamientos; en la segunda, con una media diferente; en la tercera, con dos medias diferentes; en la cuarta, con tres medias diferentes; y en la última, con todas las medias diferentes entre sí. Es importante mencionar que la mayor diferencia de medias entre tratamientos fue incorporada en la segunda etapa (es decir, cuando hay solo una media diferente), de acuerdo con la detectabilidad definida para cada caso de distribución diferente.

Para mantener la detectabilidad de las pruebas constante al agregar medias diferentes a los datos simulados, se agregaron medias que estuvieran entre la media menor y la media mayor, cuya diferencia está definida por la detectabilidad. Para ello, la mayor diferencia entre las medias simuladas, determinada por la detectabilidad, fue añadida desde la etapa en la que solo una de las medias es diferente; todas las medias diferentes que fueron añadidas en las siguientes etapas tuvieron diferencias menores para todos los casos. Asimismo, tras ser añadidas en cada escenario, las medias diferentes ya añadidas se mantuvieron para las siguientes etapas (ver Tablas 6 a 10). Se decidió proceder de esta manera con el fin de encontrar el efecto que tiene, en la potencia, el incorporar más medias diferentes entre tratamientos, poniendo como referencia al caso en el que solo una de las medias es diferente, es decir, manteniendo la detectabilidad de la prueba constante para cada caso de número de medias diferentes. La detectabilidad (δ) de cada prueba fue fijada de acuerdo a la tabla 1.

Tabla 1.

Detectabilidad definida para la potencia de la prueba por caso de distribución.

| | <u>Distribución normal</u> | <u>Distribución exponencial</u> | <u>Escala 0-4</u> | <u>Escala 0-9</u> |
|-------------------------|----------------------------|---------------------------------|-------------------|-------------------|
| Detectabilidad δ | 10 | 60 | 1.5 | 1.5 |

En cada iteración, se generaron las distribuciones de los cuatro casos de distribuciones propuestos, de acuerdo con el número de medias diferentes entre sí definidos en cada etapa. Posteriormente, se realizó una prueba de hipótesis de igualdad de medias con cinco métodos diferentes: análisis de varianza, prueba de Kruskal Wallis, prueba de permutaciones usando como estadístico de comparación al estadístico de Kruskal Wallis, prueba de permutaciones usando como estadístico de comparación al estadístico F obtenido del análisis de varianza y una prueba de permutaciones usando como estadístico de comparación al promedio de las diferencias absolutas entre cada par de medias. Para cada prueba de permutaciones se hicieron mil iteraciones; en cada una, se rechazó la hipótesis nula cada vez que el estadístico de comparación simulado con permutaciones fuera mayor al estadístico de comparación observado en cada conjunto de datos simulado para cada distribución analizada, a partir del cual se generaron las permutaciones.

Finalmente, se obtuvo la potencia de cada prueba obteniendo la proporción de veces que se rechazó la hipótesis nula de igualdad de medias, fijando una significancia de 0.05.

Como resultado de este proceso, finalmente, se obtuvo, para cada caso de distribución propuesto, una potencia para cada uno de los métodos propuestos por cada número de diferencia de medias entre los tratamientos. Para el análisis de la probabilidad de cometer error tipo I, se observó la potencia de cada método para el escenario en el que no había ninguna media diferente a las demás, es decir, donde la hipótesis nula es verdadera.

Para la simulación de los datos y los análisis descritos se utilizó el software estadístico R versión 4.3.0 (R Core Team, 2023), en el entorno de desarrollo integrado Rstudio.

RESULTADOS

Para el caso de la distribución normal, se encontró que todos los métodos analizados tienen una probabilidad de cometer error tipo I casi idéntica (cercana a 0.04) y que es menor a la significancia fijada en 0.05, por lo que, al rechazar la hipótesis nula de igualdad de medias con alguno de los tres métodos, bajo los escenarios simulados, se comprobó que no se está superando la máxima probabilidad de rechazar la hipótesis nula cuando las medias sí son iguales, es decir, que se toma una buena decisión cuando se rechaza una hipótesis nula.

Con respecto al comportamiento de la potencia cuando se agregan más medias diferentes, manteniendo la detectabilidad constante, el comportamiento general para todos los métodos propuestos fue una disminución de la potencia de la prueba cada vez que se agregaba una media diferente adicional como se puede observar en la tabla 2, por lo que, mientras más medias sean diferentes entre sí, mayor será la dificultad de que la prueba rechace la hipótesis nula de igualdad de medias.

Finalmente, la prueba más potente fue el análisis de varianza, obteniendo la mayor potencia para cada caso de número de diferencia de medias. Sin embargo, las potencias de todos los demás métodos son muy similares a las del análisis de varianza, con excepción de la prueba de permutaciones que usa como estadístico de referencia al estadístico F, que presentó potencias muy bajas en relación a los demás métodos.

Tabla 2.

Potencia de la prueba para ANOVA, Prueba de Kruskal Wallis y Permutaciones con distintos estadísticos con una distribución normal.

| <u>Diferencias</u> | <u>Anova</u> | <u>Prueba</u> <u>Kruskal Wallis</u> | <u>Permutaciones</u> <u>Estadístico F</u> | <u>Permutaciones</u> <u>Estadístico de</u> <u>Kruskal-Wallis</u> | <u>Diferencia absoluta</u> <u>promedio</u> |
|--------------------|--------------|--|--|--|---|
| 0 | 0.040 | 0.029 | 0.006 | 0.038 | 0.041 |
| 1 | 0.560 | 0.521 | 0.272 | 0.518 | 0.500 |
| 2 | 0.533 | 0.517 | 0.258 | 0.528 | 0.538 |
| 3 | 0.495 | 0.442 | 0.231 | 0.479 | 0.479 |
| Todas | 0.455 | 0.403 | 0.192 | 0.423 | 0.444 |

Para el caso de la distribución exponencial, se encontró que el análisis de varianza y la prueba de permutaciones que usa como estadístico de comparación a la diferencia absoluta promedio generaron probabilidades de cometer error tipo I mayores a la significancia fijada (0.05). Se esperaba

que el análisis de varianza produjera una probabilidad de cometer error tipo I mayor a la significancia fijada, pues se sabe que el utilizar ANOVA cuando se incumplen sus supuestos altera la significancia real de las pruebas; además, se sabe que las distribuciones exponenciales son muy asimétricas, por lo que sí había violaciones importantes al supuesto de normalidad. Por otro lado, sobre el efecto de agregar más medias diferentes, manteniendo fija la detectabilidad se puede observar en la tabla 3.

Se observó que se reduce de manera importante la potencia de las pruebas para todos los métodos analizados, bajando la probabilidad de que la prueba rechace la hipótesis nula de igualdad de medias. Finalmente, se encontró que el análisis de varianza es más potente que los demás métodos cuando el número de diferencias es menor, aunque esta ventaja es muy reducida sobre la prueba de permutaciones que usa como estadístico de comparación a la diferencia absoluta promedio. Además, esta última prueba es más potente que el resto de los métodos cuando mayor es el número de diferencias entre medias.

Tabla 3.

Potencia de la prueba para ANOVA, Prueba de Kruskal Wallis y Permutaciones con distintos estadísticos con una distribución exponencial.

| <u>Diferencias</u> | <u>Anova</u> | <u>Prueba</u> <u>Kruskal Wallis</u> | <u>Permutaciones</u> <u>Estadístico F</u> | <u>Permutaciones</u> <u>Estadístico de</u> <u>Kruskal-Wallis</u> | <u>Diferencia absoluta</u> <u>promedio</u> |
|--------------------|--------------|--|--|--|---|
| 0 | 0.060 | 0.046 | 0.006 | 0.056 | 0.064 |
| 1 | 0.322 | 0.187 | 0.137 | 0.234 | 0.319 |
| 2 | 0.302 | 0.206 | 0.105 | 0.233 | 0.331 |
| 3 | 0.211 | 0.162 | 0.075 | 0.205 | 0.244 |
| Todas | 0.156 | 0.143 | 0.045 | 0.188 | 0.198 |

En cuanto a las potencias obtenidas para la escala 0-4, se determinó que todos los métodos tienen una probabilidad de cometer error tipo I menor a la significancia fijada, por lo que se espera tomar una decisión correcta cuando se rechaza la hipótesis nula de igualdad de medias. Por lo tanto, no se encontró que se estuviera alterando la verdadera significancia de las pruebas con ninguno de los métodos, incluyendo el análisis de varianza, a pesar de que la distribución de la escala no fuera normal. Asimismo, se encontró que el agregar más medias diferentes entre tratamientos, manteniendo la detectabilidad constante, produjo una reducción en la potencia de la prueba (ver Tabla 4). Sin embargo, es relevante mencionar que esta reducción fue menor para la prueba de Kruskal Wallis y la prueba de permutaciones que usa como estadístico de comparación al estadístico de Kruskal Wallis, en relación con los demás métodos puestos a prueba.

Finalmente, se determinó que el método más potente es el análisis de varianza, seguido cercanamente por la prueba de permutaciones que utiliza como estadístico de comparación a la diferencia absoluta promedio. Este comportamiento era esperado a pesar de la violación al supuesto de normalidad, debido a que las escalas simuladas fueron parametrizadas de manera que fueran simétricas y que no tuvieran colas, por lo que se comprobó la robustez de la prueba a la violación del supuesto de normalidad cuando la distribución no es muy asimétrica ni presenta colas largas.

Tabla 4.

Potencia de la prueba para ANOVA, Prueba de Kruskal Wallis y Permutaciones con distintos estadísticos con una distribución escala de 0 a 4.

| <u>Diferencias</u> | <u>Anova</u> | <u>Prueba</u> <u>Kruskal Wallis</u> | <u>Permutaciones</u> <u>Estadístico F</u> | <u>Permutaciones</u> <u>Estadístico de</u> <u>Kruskal-Wallis</u> | <u>Diferencia absoluta</u> <u>promedio</u> |
|--------------------|--------------|--|--|--|---|
| 0 | 0.030 | 0.031 | 0.025 | 0.031 | 0.026 |
| 1 | 0.994 | 0.984 | 0.929 | 0.993 | 0.928 |
| 2 | 0.889 | 0.941 | 0.653 | 0.977 | 0.854 |
| 3 | 0.898 | 0.954 | 0.647 | 0.965 | 0.895 |
| Todas | 0.537 | 0.733 | 0.292 | 0.782 | 0.537 |

Para el caso de la escala 0-9, para todos los métodos se observó una probabilidad de cometer error tipo I menor a la significancia fijada. Con respecto a cómo afecta el añadir más medias diferentes entre tratamientos se observó que, a diferencia de los demás casos ya analizados, la potencia aumenta a medida que se añaden más medias diferentes (ver Tabla 5). Este efecto es particularmente notable para la prueba Kruskal Wallis y la prueba de permutaciones que utiliza como estadístico de comparación al estadístico de Kruskal Wallis; para ambos métodos la potencia se hizo más de seis veces mayor cuando todas las medias son diferentes que cuando sólo una de ellas lo es. Para los escenarios con menor número de medias diferentes, el análisis de varianza y la prueba de permutaciones que usa como estadístico de comparación a la diferencia absoluta promedio son más potentes que los otros métodos. Sin embargo, con mayor número de diferencias entre medias, la prueba de Kruskal Wallis y la prueba de permutaciones con el estadístico de Kruskal Wallis son mucho más potentes que los demás métodos.

Tabla 5.

Potencia de la prueba para ANOVA, Prueba de Kruskal Wallis y Permutaciones con distintos estadísticos con una distribución escala de 0 a 9.

| <u>Diferencias</u> | <u>Anova</u> | <u>Prueba</u> <u>Kruskal Wallis</u> | <u>Permutaciones</u> <u>estadístico F</u> | <u>Permutaciones</u> <u>estadístico de</u> <u>Kruskal-Wallis</u> | <u>Diferencia</u> <u>absoluta promedio</u> |
|--------------------|--------------|--|--|--|---|
| 0 | 0.049 | 0.043 | 0.007 | 0.042 | 0.049 |
| 1 | 0.209 | 0.139 | 0.039 | 0.174 | 0.222 |
| 2 | 0.214 | 0.143 | 0.042 | 0.157 | 0.212 |
| 3 | 0.217 | 0.819 | 0.036 | 0.821 | 0.222 |
| Todas | 0.253 | 0.907 | 0.073 | 0.904 | 0.249 |

CONCLUSIONES

En general, se encontró que, ante violaciones a los supuestos de normalidad y homocedasticidad, se altera la significancia real de las pruebas llevadas a cabo con el método de análisis de varianza paramétrico. Además, se determinó que para casos donde se viola el supuesto de normalidad, pero en el que la distribución de los datos no es muy asimétrica ni de colas muy largas, la violación al supuesto no es tan grave al hacer un análisis de varianza, demostrando la robustez de la prueba ante escenario como el descrito. A partir de esta noción, es relevante recalcar la importancia de analizar la naturaleza de los datos para evaluar qué alternativa de método es la apropiada para llevar a cabo inferencias estadísticas, a pesar de contar con nociones preliminares de violaciones de supuestos. Esto se vuelve relevante por la conveniencia de utilizar alternativas paramétricas por la

pérdida de información que genera no usarlas cuando es apropiado, pues, no solo pueden ser más potentes las pruebas de hipótesis, sino que utilizar alternativas paramétricas permite construir intervalos de confianza con métodos analíticos a partir de distribuciones conocidas, a diferencia de alternativas no paramétricas, como las pruebas de permutaciones, que no lo permiten.

Por otro lado, se encontró que el añadir más medias diferentes entre tratamientos manteniendo constante la detectabilidad de la prueba genera, por lo general, potencias más bajas. Sin embargo, no se busca generalizar esta conclusión, pues, en este análisis, se demostró que, para algunos métodos y distribuciones, la potencia más bien aumenta al añadir más medias diferentes entre tratamientos manteniendo constante la detectabilidad de las pruebas; esto último se observó específicamente en ambas escalas con la prueba de Kruskal Wallis y la prueba de permutaciones que utiliza como estadístico de comparación al estadístico de Kruskal Wallis.

Adicionalmente, a partir de los resultados obtenidos, se encontró que llevar a cabo una prueba de permutaciones utilizando como estadístico de prueba al estadístico de Kruskal Wallis genera potencias muy similares a llevar a cabo una prueba de Kruskal Wallis directamente. Sin embargo, no se pudo observar el mismo comportamiento entre la prueba de permutaciones que utiliza como estadístico de prueba al estadístico F y el análisis de varianza. En general, la prueba de permutaciones con el estadístico F demostró ser muy poco potente, por lo que no se recomendaría, a partir de los resultados obtenidos, utilizarla como método no paramétrico para pruebas de hipótesis de igualdad de medias. Por otro lado, la prueba de permutaciones que utilizaba como estadístico de comparación a la diferencia absoluta promedio obtuvo potencias relativamente altas para todos los casos, especialmente para la distribución normal, para la distribución exponencial en escenarios donde el número de medias diferentes sea mayor y en escalas en las que el número de medias diferentes es menor; por lo tanto, resulta de interés llevar a cabo posteriores análisis de pruebas de permutaciones con la diferencia absoluta promedio como estadístico de comparación para estudiar su potencial como alternativa no paramétrica para llevar a cabo pruebas de hipótesis de igualdad de medias para casos en los que se tengan factores con números altos de tratamientos.

REFERENCIAS BIBLIOGRÁFICAS

- Alvarado, R. (2020). Guía para el análisis de experimentos estadísticos usando R [Manuscrito no publicado]. Universidad de Costa Rica
- Bautista, F., & Gómez, E. (2007). *Una exploración de robustez de tres pruebas: dos de permutación y la de Mann-Whitney*. Revista Colombiana de Estadística, 30(2), 177-185.
- Larson, M. G. (2008). *Analysis of variance*. *Circulation*, 117(1), 115-121.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sawyer, S. (2009). Analysis of variance: the fundamental concepts. *Journal of Manual & Manipulative Therapy*, 17(2), 27E-38E.
- Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. *Applied mechanics and materials*, 611, 115-120.

ANEXOS

Tabla 6.

Medias de la distribución normal según número de tratamiento y número de diferencias entre promedios

| <u>Número de diferencias</u> | <u>Tratamiento 1</u> | <u>Tratamiento 2</u> | <u>Tratamiento 3</u> | <u>Tratamiento 4</u> | <u>Tratamiento 5</u> |
|------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 0 | 50 | 50 | 50 | 50 | 50 |
| 1 | 50 | 50 | 50 | 50 | 60 |
| 2 | 50 | 50 | 50 | 55 | 60 |
| 3 | 50 | 50 | 52.5 | 55 | 60 |
| Todas | 50 | 57.5 | 52.5 | 55 | 60 |

Tabla 7.

Medias de la distribución exponencial según número de tratamiento y número de diferencias entre promedios

| <u>Número de diferencias</u> | <u>Tratamiento 1</u> | <u>Tratamiento 2</u> | <u>Tratamiento 3</u> | <u>Tratamiento 4</u> | <u>Tratamiento 5</u> |
|------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 0 | 50 | 50 | 50 | 50 | 50 |
| 1 | 50 | 50 | 50 | 50 | 110 |
| 2 | 50 | 50 | 50 | 80 | 110 |
| 3 | 50 | 50 | 65 | 80 | 110 |
| Todas | 50 | 95 | 65 | 80 | 110 |

Tabla 8.

Varianzas de la distribución exponencial según número de tratamiento y número de diferencias entre promedios

| <u>Número de diferencias</u> | <u>Tratamiento 1</u> | <u>Tratamiento 2</u> | <u>Tratamiento 3</u> | <u>Tratamiento 4</u> | <u>Tratamiento 5</u> |
|------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 0 | 2500 | 2500 | 2500 | 2500 | 2500 |
| 1 | 2500 | 2500 | 2500 | 2500 | 12100 |
| 2 | 2500 | 2500 | 2500 | 6400 | 12100 |
| 3 | 2500 | 2500 | 4225 | 6400 | 12100 |
| Todas | 2500 | 9025 | 4225 | 6400 | 12100 |

Tabla 9.

Probabilidades de la escala 0-4 según número de tratamiento y número de diferencias entre promedios

| <u>Número de diferencias</u> | <u>Tratamiento 1</u> | <u>Tratamiento 2</u> | <u>Tratamiento 3</u> | <u>Tratamiento 4</u> | <u>Tratamiento 5</u> |
|------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| 0 | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) |
| 1 | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0.2,0.1,0.7,0,0) |
| 2 | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0.1,0.1, 0.0,0.8,0) | (0.2,0.1,0.7,0,0) |
| 3 | (0,0, 0.125,0.75,0.125) | (0,0, 0.125,0.75,0.125) | (0.0,1,0.8,0.1,0) | (0.1,0.1, 0.0,0.8,0) | (0.2,0.1,0.7,0,0) |
| Todas | (0,0, 0.125,0.75,0.125) | (0.15,0.15,0.7,0) | (0.0,1,0.8,0.1,0) | (0.1,0.1, 0.0,0.8,0) | (0.2,0.1,0.7,0,0) |

Tabla 10.

Probabilidades de la escala 0-9 según número de tratamiento y número de diferencias entre promedios

| Número de diferencias | Tratamiento 1 | Tratamiento 2 | Tratamiento 3 | Tratamiento 4 | Tratamiento 5 |
|-----------------------|------------------------|-----------------------------|-------------------------|------------------------|------------------------|
| 0 | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) |
| 1 | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (0,0,0,0,.2,0,0,0,.8) |
| 2 | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (0,1,0,0,.2,0,0,0,.7) | (0,0,0,0,.2,0,0,0,.8) |
| 3 | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (.1,.2,0,0,0,0,0,0,.7) | (0,1,0,0,.2,0,0,0,.7) | (0,0,0,0,.2,0,0,0,.8) |
| Todas | (.1,.2,0,0,0,0,0,0,.7) | (0,05,0,.1,0,0,0,1,.1,1,65) | (0,1,0,.1,0,0,0,0,7,.1) | (0,1,0,0,.2,0,0,0,.7) | (0,0,0,0,.2,0,0,0,.8) |

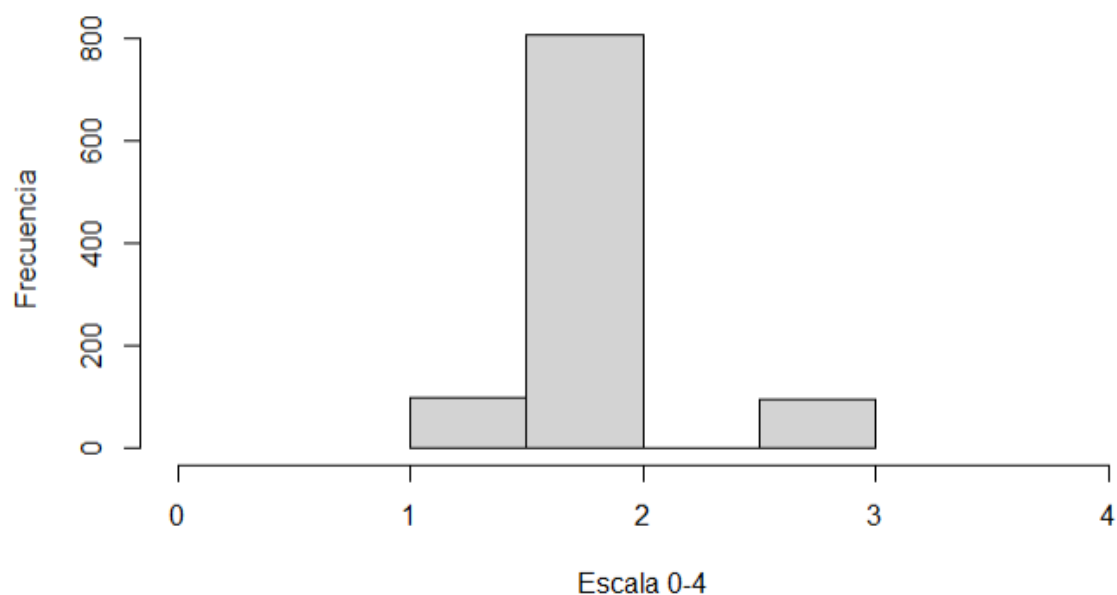


Figura 1.
Histograma de ejemplo de distribución simulada de Escala de 0 a 4.

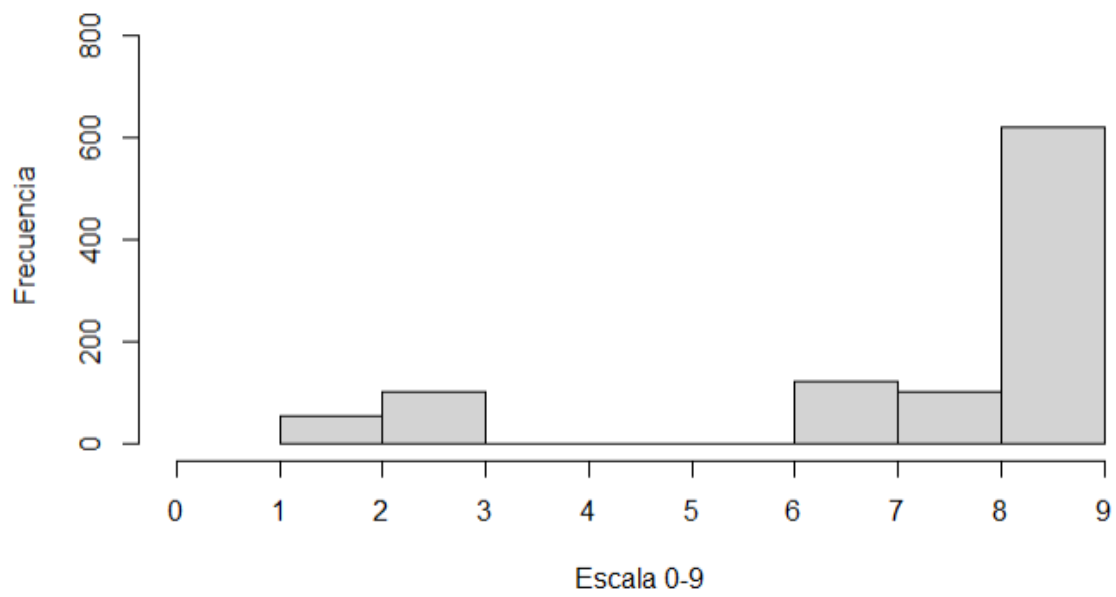


Figura 2.
Histograma de ejemplo de distribución simulada escala de 0 a 9.