

Análisis de varianza no paramétrica: un punto de vista a favor para utilizarla

Nonparametric analysis of variance: A point of view in favour to use it

Carlos A. Núñez-Colín*

RESUMEN

En este trabajo se efectuó una revisión sobre dos pruebas no paramétricas equivalentes a los modelos del análisis de varianza monofactorial de los diseños completamente al azar y bloques completos al azar, y se realizó un análisis comparativo de interpretación de ambas pruebas con sus contrapartes paramétricas. Se comparó la estadística no paramétrica empleada en muchas áreas de las ciencias con su contraparte paramétrica. Se analizó la facilidad de cálculo y la equivalencia de las pruebas Kruskal-Wallis y de Friedman a los modelos de los diseños completamente al azar y en bloques completos al azar, respectivamente, como una opción de análisis más válida cuando hay dudas o comprobaciones de que no se cumplen los supuestos del análisis de varianza.

PALABRAS CLAVE

estadística no paramétrica, prueba Kruskal-Wallis, prueba de Friedman, prueba *post hoc* de Nemenyi

ABSTRACT

The present research reviewed the use of equivalent nonparametric tests to analysis of variance in the designs completely random and complete random blocks as well as a comparative analysis between the use of both methods and its associated *post hoc* tests. We tried to explain the benefits of using nonparametric statistics in many areas of the science, trying to compare it with its parametric counterpart. We analyzed the ease of calculation and the equivalence of the Kruskal-Wallis and Friedman tests with the completely random design and complete random blocks design, respectively, as an option of more valuable analysis when there are doubts or ascertainment that the assumptions of the analysis of variance do not coincide.

KEYWORDS

nonparametric statistics, Kruskal-Wallis test, Friedman test, Nemenyi's *post hoc* test

* Programa de Ingeniería en Biotecnología, Universidad de Guanajuato. Celaya, Guanajuato, México.

Correspondencia: Mutualismo 303 esq. Prolongación Río Lerma, col. La Suiza, Apartado Postal 91. 38060, Celaya, Guanajuato, México.

Correo electrónico: lit007a@gmail.com

Recibido: 16 de marzo de 2018

Aceptado: 1 de octubre de 2018

INTRODUCCIÓN

Una de las herramientas más comunes para validar resultados en la investigación científica es el uso del análisis de varianza, principalmente cuando se requiere comparar diferentes tratamientos en experimentos. No obstante, la teoría estadística nos indica varios supuestos que se deben cumplir para aplicar dicho análisis como normalidad de la variable respuesta o también llamada variable dependiente, homogeneidad de varianzas entre los tratamientos e independencia de los datos (Infante y Zarate 2000). No siempre es posible obtener dichos datos en las investigaciones y si no son cerciorados adecuadamente, puede originarse un sesgo importante en la discusión de los resultados, por lo que pueden tomarse decisiones erróneas.

Es relativamente común en varias áreas de las ciencias no cumplir con los supuestos del análisis de varianza, sobre todo en investigaciones de sanidad vegetal donde se utilizan escalas de medición (Morales et al. 2009), en pruebas de catación de productos (Melendez et al. 2002) y de otras investigaciones subjetivas, en las cuales es complejo medir cuantitativamente la variable respuesta. Por ejemplo, cuando se está analizando una enfermedad vegetal normalmente se usa una escala de la enfermedad para categorizar los síntomas como planta sana (sin signos evidentes de la enfermedad), planta infectada (entre 20 y 50 % de daño), planta enferma (con signos evidentes de sintomatología, más de 50% de daño) y plantas muertas por la enfermedad (Placencia-Tenorio et al. 2012). Esta es una evaluación mediante variables discretas y es complicado tener normalidad, por lo que utilizar un análisis de varianza convencional, paramétrico, puede originar conclusiones erróneas.

Los análisis de varianza no paramétricos no siguen supuestos del análisis de varianza y pueden ser más flexibles y robustos que el análisis de varianza paramétrico. Sin embargo, se asume que los análisis no paramétricos son complicados y se necesita de mayor experiencia para interpretarlos, comparado con el análisis de varianza convencional. El objetivo de este trabajo fue efectuar una revisión sobre dos pruebas no paramétricas equivalentes a los modelos del análisis de varianza monofactorial de los diseños completamente al azar y bloques completos al azar y realizar un análisis comparativo de interpretación de ambas pruebas con sus contrapartes paramétricas.

MARCO DE REFERENCIA

Análisis de varianza paramétrico

El análisis de varianza en sentido estricto es describir el método estadístico desarrollado por Fisher (1937), que es utilizado para comparar si dos o más medias paramétricas difieren entre sí (Sokal y Rohlf 2002). Para el caso de dos medias se utiliza tradicionalmente la prueba de t . Sin embargo, el análisis de varianza es una prueba más general que permite contrastar dos o muchas muestras (Sokal y Rohlf 2002).

Para explicar el análisis de varianza se describen a continuación los diseños completamente al azar y bloques completos al azar.

Diseño experimental completamente al azar. El diseño experimental completamente al azar es el modelo más básico utilizado en el análisis de varianza de manera tradicional, contempla una fuente de variación de tratamientos con m unidades experimentales y el cual tiene un modelo

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}; j = 1, \dots, n_i; i = 1, \dots, t \quad 1$$

donde Y_{ij} es la respuesta en la j -ésima unidad experimental con el tratamiento i -ésimo μ es la media general, común a todas las unidades antes de aplicar los tratamientos τ_i es el efecto del i -ésimo tratamiento y ε_{ij} es el error experimental en la j -ésima unidad experimental con el tratamiento i -ésimo.

Este modelo (1) implica que mientras que las medias de las t poblaciones o tratamientos que se van a comparar son posiblemente distintas, el hecho de que todas las ε_{ij} tengan una varianza en común equivale a que los t tratamientos tienen la misma varianza, es decir, que se empieza con unidades experimentales con la misma capacidad de respuesta (μ) y con la misma varianza (σ^2). Por lo tanto, la aplicación de los tratamientos tiene el efecto de alterar las medias que ahora son $\mu_i = \mu + \tau_i$ aunque se supone que las varianzas no se modifican (Infante y Zarate 2000).

Esta suposición es después de la independencia de las ε_{ij} que es capital, la más importante, ya que el procedimiento de este análisis no es robusto, en lo más mínimo, a la violación de esta suposición. Desafortunadamente, en muchos casos, cuando la media crece ($\tau_i > 0$), la varianza también aumenta y las conclusiones que se obtienen analizando el modelo (1) son totalmente erróneas (Infante y Zarate 2000).

Cuadro 1. Tabla típica del análisis de varianza en el modelo que representa el diseño completamente al azar (Infante y Zarate 2000)

FUENTE DE VARIACIÓN	GRADOS DE LIBERTAD (G. L.)	SUMA DE CUADRADOS (S. C.)	CUADRADO MEDIO (C. M.)	F ₀
Tratamientos	$t - 1$	$\sum_{i=1}^t n_i (\bar{Y}_i - \bar{Y}_{..})^2$	$\frac{S. C. Trat}{G. L. Trat}$	$\frac{S. M. Trat}{C. M. Error}$
Error	$\sum_{i=1}^t n_i - t$	$\sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$	$\frac{S. C. Error}{S. C. Trat}$	
Total	$\sum_{i=1}^t n_i - 1$	$\sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$		

A partir de esto se obtiene la variación total del experimento y el procedimiento es hacer una partición *ad hoc* de esta variabilidad mediante la siguiente ecuación de suma de cuadrados.

$$\sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^t n_i (\bar{Y}_i - \bar{Y}_{..})^2$$

donde $\sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$ es la suma de cuadrados total (SCTot)

$\sum_{i=1}^t \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$ es la suma de cuadrados del error (SCE) y

$\sum_{i=1}^t n_i (\bar{Y}_i - \bar{Y}_{..})^2$ es la suma de cuadrados de los tratamientos (SCTrat)

Con estos datos y los grados de libertad de cada término se obtiene el cuadro comúnmente reportado para el análisis de varianza (cuadro 1). Y donde a partir de los cuadrados medios, que es el cociente de la suma de cuadrados entre los grados de libertad, se formula la prueba mediante el estadístico F para probar las hipótesis

$H_0: \mu_1 = \mu_2 = \dots = \mu_t$ vs. H_a : al menos una μ_i es diferente al resto de las demás.

Donde se rechaza la hipótesis nula si $F_0 \geq F_{G.L. Error}^{G.L. Trat}$. Por lo que se puede ver, el único factor de variación de interés en este modelo (1) es el efecto de los tratamientos en las unidades experimentales.

Diseño en bloques completos al azar. Al igual que el caso anterior, el diseño en bloques completos al azar es un modelo del análisis de varianza, pero con la diferencia de que toma en cuenta más de una fuente de variación.

Este modelo se utiliza cuando hay un gradiente en las repeticiones de los tratamientos que impiden aleatorizar de manera efectiva las unidades experimentales, por lo que se ajusta para homogeneizar las condiciones de cada tratamiento en cada bloque, es decir, el modelo corresponde a tener sólo una repetición de cada tratamiento en cada bloque. Por lo que el modelo quedaría expuesto de la siguiente manera.

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}; j = 1, \dots, b; i = 1, \dots, t \quad 2$$

donde Y_{ij} es la respuesta en la j -ésima unidad experimental con el tratamiento i -ésimo

μ es la media general, común a todas las unidades antes de aplicar los tratamientos

τ_i es el efecto del i -ésimo tratamiento

β_j es el efecto del j -ésimo bloque y

ε_{ij} es el error experimental en la j -ésima repetición del i -ésimo tratamiento.

En este modelo (2) se pondera el efecto del bloque (repetición) sobre las unidades experimentales sin considerar esta fuente de variación como de interés para la investigación sino como una variación aleatoria, ajena al planteamiento de los tratamientos del experimento pero que puede afectar la visualización de los mismos. Esto es utilizado para probar los tratamientos en condiciones diversas (por eso se hacen los bloques). Además, se presupone que no hay interacción entre los bloques y los tratamientos (Sokal y Rohlf 2002) y todos los tratamientos deben estar representados obligatoriamente una vez en cada

Cuadro 2. Tabla típica del análisis de varianza en el modelo que representa el diseño en bloques completos al azar (Infante y Zarate 2000)

FUENTE DE VARIACIÓN	GRADOS DE LIBERTAD (G. L.)	SUMA DE CUADRADOS (S. C.)	CUADRADO MEDIO (C. M.)	F ₀
Tratamientos	$t - 1$	$b \sum_{i=1}^t (\bar{Y}_i - \bar{Y}_{..})^2$	$\frac{S. C. Trat}{G. L. Trat}$	$\frac{C. M. Trat}{C. M. Error}$
Bloques	$b - 1$	$t \sum_{j=1}^b (\bar{Y}_{.j} - \bar{Y}_{..})^2$	$\frac{S. C. Bloq}{G. L. Bloq}$	$\frac{C. M. Bloq}{C. M. Error}$
Error	$(b - 1)(t - 1)$	$\sum_{i=1}^t \sum_{j=1}^b (Y_{ij} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}_{..})^2$	$\frac{S. C. Error}{G. L. Error}$	
Total	$bt - 1$	$\sum_{i=1}^t \sum_{j=1}^b (Y_{ij} - \bar{Y}_{..})^2$		

bloque, por lo que todos los tratamientos deben tener el mismo número de repeticiones. Al igual que el caso anterior, las suposiciones hechas con este modelo (2) son aún más estrictas que para el modelo en un diseño completamente al azar (1) y generalmente no se cumplen por lo que puede haber un sesgo importante.

A partir de este modelo (2) se obtiene la variación total del experimento y el procedimiento es hacer una partición *ad hoc* de esta variabilidad mediante la siguiente ecuación de suma de cuadrados.

$$\sum_{i=1}^t \sum_{j=1}^b (Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^t \sum_{j=1}^b (Y_{ij} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}_{..})^2 + b \sum_{i=1}^t (\bar{Y}_i - \bar{Y}_{..})^2 + t \sum_{j=1}^b (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

donde $\sum_{i=1}^t \sum_{j=1}^b (Y_{ij} - \bar{Y}_{..})^2$ es la suma de cuadrados total (SCTot)

$\sum_{i=1}^t \sum_{j=1}^b (Y_{ij} - \bar{Y}_i - \bar{Y}_{.j} + \bar{Y}_{..})^2$ es la suma de cuadrados del error (SCE)

$b \sum_{i=1}^t (\bar{Y}_i - \bar{Y}_{..})^2$ es la suma de cuadrados de los tratamientos (SCTrat) y

$t \sum_{j=1}^b (\bar{Y}_{.j} - \bar{Y}_{..})^2$ es la suma de cuadrados de bloques (SCBloq)

Con estos datos y los grados de libertad de cada término se obtiene el cuadro comúnmente reportado para el análisis de varianza (cuadro 2).

A partir de los cuadrados medios, que es el cociente de la suma de cuadrados entre los grados de libertad,

se formula la prueba mediante el estadístico F para probar las hipótesis

$H_0: \mu_1 = \mu_2 = \dots = \mu_t$ vs. H_a : al menos una μ_i es diferente al resto de las demás. Donde se rechaza

la hipótesis nula si $F_0 \geq F_{G.L. Error}^{G.L. Trat}$. Por lo que se

puede ver, el único factor de variación de interés en este modelo (2) es, al igual que el anterior, el efecto de los tratamientos en las unidades experimentales, aun cuando se puede generar una prueba de hipótesis para bloques este efecto solo debe considerarse de ponderación ya que, si no es así, debe mejor emplearse un diseño de tratamientos factorial.

Pruebas de medias (pruebas post hoc). El análisis de varianza no permite saber exactamente las diferencias entre los tratamientos para lo que comúnmente se aplica una prueba de medias a partir del análisis de varianza (sin importar su diseño). Las pruebas de medias más utilizadas son la diferencia mínima significativa (LSD, por sus siglas en inglés) que es calcular mediante la distribución *t* el valor mínimo significativo para considerar las medias como diferentes; la prueba de Tukey (1949), la cual se denominará como la prueba de la diferencia honestamente significativa (HSD, por sus siglas en inglés) y la cual es similar a la anterior pero que difiere porque Tukey modeló una tabla de probabilidades *ad hoc* (tabla *q*) a su prueba, lo que la hace más robusta que la LSD, y la prueba de rangos de Duncan (1955), que es también basada en *t* pero que considera valores diferenciales

dependiendo del número de tratamientos que se estén comparando. Sin embargo, esta última no es estable cuando se comparan muchos tratamientos, lo que complica también el cálculo de muchas diferencias mínimas significativas.

Aunque las pruebas LSD, prueba de Tukey y prueba de rangos de Duncan son válidas para hacer la prueba *post hoc* del análisis de varianza, la mayoría de los investigadores ha optado por utilizar la LSD y la HSD, por lo que a continuación se describen.

Prueba LSD. La prueba LSD es la comparación de medias por pares y es calculada mediante la fórmula $t_{G.L.E.,\alpha} \sqrt{CME \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$, donde el valor de t está relacionado a los grados de libertad del error del análisis de varianza y toma en cuenta el cuadrado medio del error del análisis de varianza como la varianza general, y donde n_i y n_j son el número de repeticiones para el cálculo de las medias i y la media j , éstas pueden ser iguales, lo que reduce la expresión a $t_{G.L.E.,\alpha} \sqrt{\frac{2CME}{n}}$, esta diferencia mínima significativa se compara con el valor absoluto de la diferencia entre ambas medias, si esta diferencia es mayor a la diferencia mínima significativa entonces se concluye que las medias son estadísticamente diferentes.

Prueba HSD. Al igual que el caso anterior, esta prueba es para hacer la comparación por pares, a diferencia de la LSD ocupa el rango studentizado q que fue hecha ex profeso para esta prueba. Por lo que la HSD es calculada mediante la fórmula $q_{\alpha,k,G.L.E} \sqrt{\frac{CME}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$, donde k es el número de tratamientos a comparar y si se tiene el mismo número de repeticiones entonces la fórmula se reduce a $q_{\alpha,k,G.L.E} \sqrt{\frac{CME}{n}}$, al igual que el caso anterior la diferencia honesta significativa se compara con el valor absoluto de la diferencia entre ambas medias, si esta diferencia es mayor a la diferencia honesta significativa entonces se concluye que las medias son estadísticamente diferentes.

Análisis de varianza no paramétrico

Cuando se sospecha que los datos a analizar no cumplen con los supuestos entonces se ven alternativas de análisis a técnicas más robustas pero que hagan lo mismo, que sean análogas, que el análisis de varianza paramétrico o incluso puede realizarse aun cuando cumplan los supuestos. Sin embargo, Sokal y Rohlf (2002, p. 204) mencionan "...[las] técnicas (no paramétricas) se utilizan a veces por preferencia aun cuando el método paramétrico (análisis de varianza en este caso) pueda emplearse legítimamente. La

rapidez del cálculo y una preferencia de los supuestos generalmente sencillos de los análisis no paramétricos hace que muchos investigadores recurran a ellas. No obstante, cuando se reúnen los requisitos de análisis de la varianza estos métodos son menos eficientes que él". En este sentido, se concuerda con estos autores que, si se reúnen los supuestos, la primera opción debe ser el análisis paramétrico, pero si hay dudas o hay una comprobación de que no cumple algún supuesto es mejor emplear un análisis no paramétrico.

Las técnicas no paramétricas análogas al análisis de varianza de manera general se presuponen se comparan las medianas en lugar de la media (Townend 2005; Daniel 2005). No obstante, sus cálculos se hacen a través de la transformación de los datos en rangos y a partir de este se genera el estadístico de prueba principalmente basado en la distribución χ^2 (Siegel y Castellan 2005). En este sentido, tanto el promedio como la mediana son estimadores de la esperanza matemática de la distribución normal.

Asimismo, la técnica no paramétrica análoga al modelo completamente al azar (1) del análisis de varianza es la prueba de Kruskal-Wallis (Kruskal y Wallis 1952), que es una ampliación de la prueba de rangos de Wilcoxon (1945), la cual sería la versión no paramétrica de la prueba de comparación de medias de t . El cual es una prueba aplicada a diferentes áreas como la ciencia de la salud (Breslow 1970; Gómez-Gómez et al. 2003), ciencias de la comunicación (Ihom et al. 2011), mecánica aplicada y ciencia de los materiales (Ostertagová et al. 2014), entre otras. La cual es utilizada por el mínimo de supuestos que se necesita para aplicarla por lo que es muy versátil (Daniel 2005; Ihom et al. 2011) y se adapta mejor a datos ordinales y en escalas que pueden ser subjetivas (Siegel y Castellan 2005; Ortettagová et al. 2014).

La técnica no paramétrica análoga al modelo en bloques completos al azar (2) del análisis de varianza sería la prueba de Friedman (1940), que es de las primeras ideas sobre el uso de rangos para técnicas estadísticas análogas al análisis de varianza y que actualmente se ha adaptado para evaluar análogamente el modelo en bloques completos al azar (2). Esta prueba es aplicada en ciencias de la conducta (Siegel y Castellan 2005), ciencias de la salud (Sheldon et al. 1996), ciencias de la comunicación (Pereira et al. 2015), entre otras. Estas técnicas son más recomendadas cuando se tienen datos pareados, que servirían para hacer los bloques, por lo que el diseño debe ser más estricto que para la prueba Kruskal-Wallis (Sheldon et al. 1996; Siegel y Castellan 2005). Ambas técnicas se describen a continuación más ampliamente.

Prueba de Kruskal-Wallis. Para aplicar esta prueba se debe conocer primero el número total de unidades experimentales (N) y el número de muestras de cada tratamiento (n_i), puede que cada tratamiento no tenga el mismo número de repeticiones. Acto seguido, se ordenan y se le asigna el valor de uno al valor más bajo, dos al siguiente valor; el último será el valor más alto y tendrá el valor de N . En caso de empate, los valores (dos o más empatados con el mismo valor) se identifican los valores que les correspondería en el orden de los rangos y se promedian y se les asigna ese valor promedio a todos los datos con dicho valor. Una vez que se tienen todos los valores de los rangos, estos se agrupan por tratamiento y se suman teniendo un valor de suma de rangos por tratamiento (equivalente a la suma de cuadrados) y se aplica el siguiente estadístico de prueba

$$KW = \left[\left(\frac{12}{N(N+1)} \right) \sum_{i=1}^k \frac{R_i^2}{n_i} \right] - 3(N+1)$$

donde N es el número total de datos a evaluar
 n_i es el número de repeticiones del tratamiento i
 k es el número de tratamientos que se está comparando y
 R_i^2 es la suma de rangos al cuadrado del tratamiento i

Y donde a partir de la suma de rangos cuadrados, se formula la prueba mediante el estadístico χ^2 para probar las hipótesis $H_0: \tau_1 = \tau_2 = \dots = \tau_i$ vs. H_a : al menos una τ_i es diferente al resto de las demás. Donde se rechaza la hipótesis nula si $KW \geq \chi_{k-1, \alpha}^2$. Por lo que se puede observar, el planteamiento del análisis de varianza con el modelo (1) es análogo a esta prueba y se trata ver sólo el efecto de los tratamientos en las unidades experimentales. No obstante, una de las críticas más comunes es la falta de pruebas *post hoc* para esta prueba, por lo que se incluye la prueba usando la distribución normal estándar (Z) (Siegel y Castellan 2005) y la prueba de Nemenyi (1963) que usa el rango studentizado q para su cálculo.

Pruebas post hoc de la prueba Kruskal-Wallis. La primera prueba se calcula a partir de los rangos medios ($\bar{R}_i = \frac{R_i}{n_i}$) mediante la fórmula $Z_{\alpha/k(k-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$. Esta fórmula se puede reducir si se tienen el mismo número de repeticiones a la siguiente expresión $Z_{\alpha/k(k-1)} \sqrt{\frac{k(N+1)}{6}}$. Al igual que en las pruebas *post hoc* paramétricas, se compara con el valor absoluto de la diferencia entre ambos rangos medios, si esta diferencia es mayor a la diferencia mínima significativa entonces se concluye que los tratamientos son estadísticamente diferentes.

La prueba de Nemenyi (1963) puede hacerse para rangos medios o para suma de rangos, aquí sólo se expondrá la fórmula para rangos medios la cual se obtiene mediante la siguiente fórmula $q_{\alpha, \infty, k} \sqrt{\frac{N(N+1)}{24} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$ y la cual puede reducirse si se tienen el mismo número de repeticiones por tratamiento a $q_{\alpha, \infty, k} \sqrt{\frac{k(N+1)}{12}}$. Al igual que en la anterior, se compara con el valor absoluto de la diferencia entre ambos rangos medios, si esta diferencia es mayor a la diferencia honesta significativa, entonces se concluye que los tratamientos son estadísticamente diferentes.

Prueba de Friedman. A diferencia de la prueba de Kruskal-Wallis, la Prueba de Friedman (1940) considera una fuente de variación externa, los bloques, y por lo tanto los rangos se calculan diferente, ya que los rangos no se hacen generales, sino que se clasifican los rangos dentro de cada uno de los bloques. Por ejemplo, si se tienen cuatro tratamientos y tres bloques en cada uno de esos tres bloques los rangos tendrán valores del uno al cuatro y si hay empates dentro del bloque se sigue el mismo procedimiento descrito en la prueba Kruskal-Wallis, pero se considera el mismo supuesto que se debe tener una repetición por tratamiento en cada bloque, es decir, todos los tratamientos tienen el mismo número de repeticiones. Una vez que se tienen los rangos se agrupan por tratamiento y se suman teniendo un valor se suma de rangos (equivalente a la suma de cuadrados) y se aplica el siguiente estadístico de prueba

$$Fr = \left[\left(\frac{12}{bk(k+1)} \right) \sum_{i=1}^k R_i^2 \right] - 3b(k+1)$$

donde b es el número de bloques
 k es el número de tratamientos que se está comparando y
 R_i^2 es la suma de rangos al cuadrado del tratamiento i

Y donde a partir de la suma de rangos cuadrados, se formula la prueba mediante el estadístico χ^2 para probar las hipótesis $H_0: \tau_1 = \tau_2 = \dots = \tau_i$ vs. H_a : al menos una τ_i es diferente al resto de las demás. Donde se rechaza la hipótesis nula si $Fr \geq \chi_{k-1, \alpha}^2$. Por lo que se puede observar, el planteamiento del análisis de varianza con el modelo (2) es análogo a esta prueba y se trata ver sólo el efecto de los tratamientos en las unidades experimentales, no de los bloques ya que este efecto solo debe considerarse de ponderación para generalizar el efecto de los tratamientos en diversas condiciones. Por ejemplo, en una catación de vinos los bloques son las personas que prueban

Cuadro 3. Datos del número de semillas por fruto en diez frutos por árbol de cuatro variedades de chirimoya (*Annona cherimola* Mill.) (adaptado de Andrés-Agustín 2004).

	VARIEDAD 1		VARIEDAD 2		VARIEDAD 3		VARIEDAD 4	
	Número de semillas	Rango Wilcoxon						
Fruto 1	23	3.5	132	38	106	32	46	11
Fruto 2	51	13	49	12	75	23.5	36	8
Fruto 3	26	5	172	40	59	17.5	23	3.5
Fruto 4	29	6	129	36	82	28	62	20
Fruto 5	37	9	55	14	57	15.5	83	29
Fruto 6	18	2	57	15.5	128	35	77	25.5
Fruto 7	66	21	131	37	75	23.5	16	1
Fruto 8	33	7	115	34	61	19	94	30
Fruto 9	77	25.5	108	33	101	31	79	27
Fruto 10	59	17.5	135	39	70	22	44	10
Promedio	41.9		108.3		81.4		56.0	
Mediana	35.0		122.0		75.0		54.0	
Desviación estándar	20.1		41.2		23.3		26.9	
Varianza	404.3		1701.1		542.9		723.6	

los distintos vinos, pero lo que interesa es qué vino es mejor no catalogar a los catadores (González 2015), por lo que esta prueba es *ad hoc* para evaluar este tipo de estudios.

Pruebas post hoc de la prueba Friedman. Al igual que con la prueba Kruskal-Wallis, se expondrán dos versiones de la prueba *post hoc* que tienen el mismo principio e interpretación. La primera prueba es la de Z (Siegel y Castellan 2005), que se calcula a partir de la suma de rangos mediante la fórmula $Z_{\alpha/k(k-1)} \sqrt{\frac{bk(k+1)}{6}}$. Al igual que en las pruebas *post hoc* paramétricas se compara con el valor absoluto de la diferencia entre ambas sumas de rangos, si esta diferencia es mayor a la diferencia mínima significativa, entonces se concluye que los tratamientos son estadísticamente diferentes. La segunda es la versión de Nemenyi (1963) esta prueba es aplicada a los rangos medios ($\bar{R}_i = \frac{R_i}{b}$) y se calcula por la fórmula $\frac{q_{\alpha,\infty,k}}{\sqrt{2}} \sqrt{\frac{k(k+1)}{6b}}$. Al igual que en la anterior, se compara con el valor absoluto de la diferencia entre ambos rangos medios, si esta diferencia es mayor a la diferencia honesta significativa entonces se concluye que los tratamientos son estadísticamente diferentes.

EJEMPLOS DE USO

Para ejemplificar el uso de estadística no paramétrica y en particular de estas pruebas se usarán dos ejemplos de Andrés-Agustín (2004), el primero para semillas por fruto de chirimoya (*Annona cherimola* Mill.) en cuatro variedades, un árbol por variedad, y el segundo con el peso de frutos por árbol de tres variedades con un gradiente de edad de los árboles. Para el primer ejemplo se tienen en cuenta los datos del cuadro 3.

Las varianzas en cada variedad son heterogéneas por lo que utilizar un análisis de varianza paramétrico pudiera generar sesgo por no cumplir con este supuesto. Al considerarlo en un experimento completamente al azar se aplica una prueba de Kruskal-Wallis, detallada anteriormente. Se ordenan de menor a mayor y se les asigna un rango Wilcoxon. En los valores iguales se promedian los rangos que se les asignaría para que ambos datos tengan el mismo valor (cuadro 3). Donde se obtienen los siguientes datos para alimentar a la fórmula de Kruskal-Wallis: R1 = 109.5 (suma de rangos de la variedad 1); R2 = 298.5 (suma de rangos de la variedad 2); R3 = 247 (suma de

rangos de la variedad 3); R4 = 165 (suma de rangos de la variedad 4); k = número De tratamientos= 4. n1 = n2 = n3 = n4= 10 (tamaño de muestra de cada variedad); N = 40 (número total de observaciones). Por lo que el valor de KW y χ^2_t es calculado de la siguiente forma:

$$KW = \left[\frac{12}{N(N+1)} \cdot \sum_{j=1}^k \frac{R_j^2}{n_j} \right] - 3(N+1) \\ = \left[\frac{12}{40(41)} \cdot \left(\frac{109.5^2}{10} + \frac{298.5^2}{10} + \frac{247^2}{10} + \frac{165^2}{10} \right) \right] - 3(41) = 15.53$$

$$\chi^2_t = \chi^2_{k-1, \alpha} = \chi^2_{3, 0.05} = 7.81$$

Y como $KW > \chi^2_t$ entonces se rechaza la hipótesis nula, teniendo que al menos una variedad es diferente a las demás, para poder concluir de buena forma se procede a aplicar una prueba *post hoc*, en este caso se probó la prueba de Nemenyi (1963) para rangos medios ($\bar{R}_j = \frac{R_j}{n_j}$).

Donde la diferencia honesta significativa se estimó mediante la siguiente expresión.

$$DHS = q_{\alpha, \infty, k} \sqrt{\frac{k(N+1)}{12}} = q_{0.05, \infty, 4} \sqrt{\frac{4(40+1)}{12}} = (3.63) \sqrt{\frac{4(41)}{12}} = 13.42$$

Y por lo tanto se tienen los siguientes cálculos

|R1 – R2| = 18.90 > 13.42 por lo tanto V1 ≠ V2
 |R1 – R3| = 13.75 > 13.42 por lo tanto V1 ≠ V3
 |R1 – R4| = 05.55 < 13.42 por lo tanto V1 = V4
 |R2 – R3| = 05.15 < 13.42 por lo tanto V2 = V3
 |R2 – R4| = 13.35 < 13.42 por lo tanto V2 = V4
 |R3 – R4| = 08.20 < 13.42 por lo tanto V3 = V4

Y se puede establecer como normalmente se presentan este tipo de pruebas (cuadro 4). Aunque hay que recordar que la prueba está hecha en rangos medios, estos valores sólo sirven para asignar las literales de la prueba, pero es mejor acompañarlas de su promedio o su mediana para mejor entendimiento. Lo que genera que la presentación de los datos sea exactamente el mismo que para pruebas *post hoc* de análisis de varianza paramétrico.

Para el segundo se toma en cuenta otra fuente de variación por lo que se considera que está en un diseño en bloques al azar, pero en este caso también presentan varianzas que probablemente sean heterogéneas que es uno de los supuestos del análisis de varianza paramétrico (cuadro 5), por lo que se procede a aplicar la prueba de Friedman. Los datos a procesar se concentran en el cuadro 5.

Cuadro 4. Presentación de la prueba de Nemenyi ($\alpha=0.05$) del número de semillas por fruto en diez frutos por árbol de cuatro variedades de chirimoya (*Annona cherimola* Mill.) (adaptado de Andrés-Agustín 2004).

VARIEDAD	PROMEDIO DEL NÚMERO DE SEMILLAS POR FRUTO	MEDIANA DEL NÚMERO DE SEMILLAS POR FRUTO
Variiedad 2	108.3 a	122 a
Variiedad 3	81.4 a	75 a
Variiedad 4	56.0 ab	54 ab
Variiedad 1	41.9 b	35 b

Letras diferentes representan diferencias significativas mediante la prueba de Nemenyi ($\alpha=0.05$)

A diferencia del anterior, los rangos se hacen para cada bloque, en este caso el bloque representa la edad del árbol. Teniendo que R1 =26 (rango de la variedad 1); R2 =15 (rango de la variedad 2); R3 =19 (rango de la variedad 3); b=10 (número de bloques) y k=3 (número de tratamientos). El valor de χ^2_c y χ^2_t es calculado de la siguiente forma:

$$Fr = \left(\frac{12}{bk(k-1)} \sum_{i=1}^k R_i^2 \right) - 3b(k+1) \\ = \left(\frac{12}{10(3)(3-1)} \cdot (26^2 + 15^2 + 19^2) \right) - 3(10)(3+1) = 6.2$$

$$\chi^2_t = \chi^2_{k-1, \alpha} = \chi^2_{2, 0.05} = 5.99$$

Y como $Fr > \chi^2_t$ entonces se rechaza la hipótesis nula, teniendo que al menos una variedad es diferente a las demás, para poder concluir de buena forma se procede a aplicar una prueba *post hoc*, en este caso se probó la prueba de Nemenyi (1963) para rangos medios ($\bar{R}_j = \frac{R_j}{n_j}$).

Donde la diferencia honesta significativa se estimó mediante la siguiente expresión.

$$DHS = \frac{q_{\alpha, \infty, k}}{\sqrt{2}} \sqrt{\frac{k(k+1)}{6b}} = \frac{q_{0.05, \infty, 3}}{\sqrt{2}} \sqrt{\frac{3(3+1)}{6(10)}} = \frac{3.31}{1.41} \cdot \sqrt{\frac{3(4)}{6(10)}} = 1.07$$

Y por lo tanto se tienen los siguientes cálculos

|R1 – R2| = 1.10 > 1.07 por lo tanto V1 ≠ V2
 |R1 – R3| = 0.70 < 1.07 por lo tanto V1 = V3
 |R2 – R3| = 0.40 < 1.07 por lo tanto V2 = V3

Y, al igual que en ejemplo anterior, se puede establecer como normalmente se presentan este tipo de pruebas (cuadro 6).

Cuadro 5. Datos del número de semillas por fruto en diez frutos por árbol de cuatro variedades de chirimoya (*Annona cherimola* Mill.) (adaptado de Andrés-Agustín 2004).

EDAD DE LOS ÁRBOLES (BLOQUES)	VARIABLE	VARIEDAD 1	VARIEDAD 2	VARIEDAD 3
un año	peso de frutos	855	760	999
	Rango de Friedman	2	1	3
un año y medio	peso de frutos	1113	504	645
	Rango de Friedman	3	1	2
dos años	peso de frutos	719	412	880
	Rango de Friedman	2	1	3
dos años y medio	peso de frutos	924	673	1245
	Rango de Friedman	2	1	3
tres años	peso de frutos	1115	808	883
	Rango de Friedman	3	1	2
tres años y medio	peso de frutos	921	764	422
	Rango de Friedman	3	2	1
cuatro años	peso de frutos	1106	356	970
	Rango de Friedman	3	1	2
cuatro años y medio	peso de frutos	912	951	416
	Rango de Friedman	2	3	1
cinco años	peso de frutos	1170	856	143
	Rango de Friedman	3	2	1
más de cinco años	peso de frutos	1190	560	118
	Rango de Friedman	3	2	1
	Promedio	1002.5	664.4	672.1
	Mediana	1015.0	762.0	762.5
	Desviación estándar	157.0	198.3	384.0
	Varianza	24657.2	39338.7	147485.4

Lo que genera, al igual que el ejemplo anterior, que la presentación de los datos sea exactamente el mismo que para pruebas *post hoc* de análisis de varianza paramétrico.

DISCUSIÓN

El método no paramétrico de Kruskal-Wallis es análogo al análisis de varianza con un diseño completamente al azar, y el de Friedman, al análisis de varianza con un diseño en bloques completos al azar. Aunque sus cálculos son diferentes, en ambos métodos no paramétricos estos son más sencillos que sus versiones paramétricas, incluso para sus pruebas *post hoc*. Lo que hace que la interpretación de la prueba no paramétrica sea exactamente igual que su contraparte paramétrica.

Siegel y Castellan (2005) mencionan que los métodos no paramétricos se centran en el orden o el rango de las unidades experimentales más que en sus valores numéricos. Ihom et al. (2011) demuestra que los resultados de la prueba de Kruskal-Wallis es idéntico al que se obtiene con un análisis de varianza de una vía (modelo 1). No obstante, la principal justificación de estas técnicas no paramétricas es que los supuestos que requiere el análisis de varianza paramétrico ocasionan que en muchos de los casos experimentales estos no se cumplan. Como resultado puede haber sesgo y la no repetitividad de los resultados, lo que dificulta generalizarlos. La mayoría de los profesores de estadística enseña los métodos paramétricos y son pocos los que también enseñan y recalcan la utilidad de los métodos no paramétricos, los cuales deberían ser parte de la formación académica en carreras de

Cuadro 6. Presentación de la prueba de Nemenyi ($\alpha=0.05$) del peso de frutos por árbol en tres variedades de chirimoya (*Annona cherimola* Mill.) (adaptado de Andrés-Agustín 2004).

VARIEDAD	PROMEDIO DEL PESO DE FRUTOS POR ÁRBOL	MEDIANA DEL PESO DE FRUTO POR ÁRBOL
Variedad 1	1002.5 a	1015.0 a
Variedad 3	672.1 ab	762.5 ab
Variedad 2	664.4 b	762.0 b

Letras diferentes representan diferencias significativas mediante la prueba de Nemenyi ($\alpha=0.05$)

ciencias de la salud, ciencia de los alimentos y sanidad vegetal y animal. En estas áreas la mayoría de los resultados son en escalas no normales (porcentajes, escalas de daño, catas, valoraciones médicas cualitativas) y en ocasiones utilizan muchos ceros que hace que las varianzas de los tratamientos no sean equiparables y por lo tanto heterogéneas.

Autores como Sokal y Rohlf (2002), Daniel (2005) y Little y Hills (2008) recomiendan transformar datos para tratar de corregir el problema, principalmente mediante raíz cuadrada o logaritmo y la transformación angular para datos en porcentajes. No obstante, esto no siempre corrige el problema y hay que hacer pruebas adicionales como la Kolmogorov-Smirnov para cerciorarse de la normalidad y la prueba de Bartlett para homogeneidad de varianzas, además de constatar la no interacción de los bloques con los tratamientos en el modelo 2. Todo lo anterior aunado a lo fácil de cálculo y del establecimiento de los rangos hace de estos métodos no paramétricos una buena opción de análisis cuando se va a comparar tratamientos, siempre y cuando estén establecidos bajo los modelos en estudios, limitando su uso para el caso de diseño de tratamientos factoriales, donde, aunque hay avances, no hay una prueba estándar para resolver este tipo de planteamientos. Sin embargo, en muchos experimentos de ciencias de la salud, ciencia de los alimentos y sanidad vegetal y animal las variables de respuesta obtenidas en ocasiones no cumplen con los supuestos del análisis de varianza paramétrico, por lo que es muy recomendable el uso de estos métodos no paramétricos. Principalmente cuando se ponen escalas de medición, porcentajes o en datos ordinales.

COMENTARIOS FINALES

La estadística no paramétrica es ideal para datos con distribuciones no normales o distribuciones "libres" (Free distribution samples), que en el caso de ciencias biológicas, agronómicas y de la salud son comunes y la falta de criterio o inexperiencia de muchos investigadores y estudiantes hace que no hagan las pruebas preliminares para cerciorarse de cumplir con los estrictos supuestos de los análisis paramétricos, lo que ocasiona concluir cuestiones erróneas y tener contradicciones estadísticas, por ejemplo, tener diferencias significativas en el análisis de varianza, pero no encontrarlas en las pruebas *post hoc*. Todo esto favorece el uso de la estadística no paramétrica. No obstante, cuando los supuestos se cumplen, es recomendable utilizar el método paramétrico y que normalmente es más eficiente y existen mayores programas estadísticos que los calculan. Aunque ya hay programas estadísticos que calculan pruebas no paramétricas, la mayoría aún no hacen, por ejemplo, las pruebas *post hoc* que son muy importantes para un mejor entendimiento de los resultados, lo que ha limitado su uso y dominio del tema en muchas áreas.

Finalmente, el cálculo de estos métodos hace que cualquier estudiante o investigador con una calculadora pueda realizar por completo su análisis estadístico sin necesidad de programas especializados. Se debe enfatizar que los métodos no paramétricos expuestos aquí son pruebas estándar que se deben conocer y aplicar en muchas áreas de la ciencia y evitar el uso indiscriminado de los métodos paramétricos sin el estricto rigor de cerciorar que cumplen con todos los supuestos de dichas pruebas.

AGRADECIMIENTOS

El autor agradece al editor Dr. Víctor López Martínez y a los árbitros anónimos por sus atinados comentarios para enriquecer el presente trabajo. Este trabajo forma parte de los resultados del proyecto CIIC 297/2018 de la Universidad de Guanajuato.

LITERATURA CITADA

- Andrés-Agustín J. 2004. Caracterización morfológica y postcosecha de las principales selecciones y variedades de chirimoyo (*Annona cherimola* Mill) en México. Tesis de doctorado. Universidad Autónoma Chapingo, Chapingo, México.
- Breslow N. 1970. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* 57(3): 579-594. <https://doi.org/10.1093/biomet/57.3.579>
- Daniel WW. 2008. Bioestadística. Limusa-Wiley. Ciudad de México, México.
- Duncan DB. 1955. Multiple range and multiple F tests. *Biometrics* 11(1): 1-42. <https://doi.org/10.2307/3001478>
- Fisher RA. 1937. The design of experiments. Oliver & Boyd LTD. Edinburgh, UK.
- Friedman M. 1940. A comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics* 11(1): 86-92. <https://doi.org/10.1214/aoms/1177731944>
- Gómez-Gómez M, Danglot-Banck C, Vega-Franco L. 2003. Sinopsis de pruebas estadísticas no paramétricas. Cuando usarlas. *Revista Mexicana de Pediatría* 70(2): 91-99.
- González M. 2015. Características sensoriales y composición no volátil de vinos tintos: avances en la exploración de la calidad. Tesis de doctorado. Universidad de la Rioja, La Rioja, España.
- Ihom PA, Abella AS, Anbua E, Ogbodo J. 2011. Kruskal-Wallis Test as analytical tool for key components of a newly developed core mixture. *Leonardo Electronic Journal of Practices and Technologies* 10(18): 114-122.
- Infante S, Zarate GP. 2000. Métodos estadísticos, un enfoque interdisciplinario. Trillas. Ciudad de México, México.
- Kruskal WH, Wallis WA. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47(260): 583-621. <https://doi.org/10.1080/01621459.1952.10483441>
- Little TM, Hills FJ. 2008. Métodos estadísticos para la investigación en la agricultura. Trillas. Ciudad de México, México.
- Meléndez E, Gil ML, Falqué E. 2002. Potencialidad de la variedad Torrentés para elaborar vinos monovarietales. *Ciencia y Tecnología Alimentaria* 3(4): 223-228. <https://doi.org/10.1080/11358120209487732>
- Morales JL, Rodríguez MP, Azpíroz HS, Pedraza ME. 2009. Modelo para la estimación del área del fruto en la evaluación de la antracnosis en aguacate (*Persea americana* Mill.) cv. Hass. *Revista UDO Agrícola* 9(2): 421-424.
- Nemenyi PB. 1963. Distribution-free multiple comparisons. PhD thesis. Princeton University, Princeton, USA.
- Ostertagová E, Ostertag O, Kováč J. 2014. Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials* 611: 115-120. <https://doi.org/10.4028/www.scientific.net/AMM.611.115>
- Pereira DG, Alfonso A, Madeiros FM. 2015. Overview of Friedman's test and post-hoc analysis. *Communications in Statistics – Simulation and Computation* 44: 2636-2653. <https://doi.org/10.1080/03610918.2014.931971>
- Plascencia-Tenorio RI, Olalde-Portugal V, Mena-Violante HG, Ceja-Torres LF, Venegas-González J, Oyoque-Salcedo G, Angoa-Pérez MV. 2012. Antagonismo in vitro de aislados bacterianos de fresa comercial y silvestre vs. *Botrytis cinerea* y *Rhizopus stolonifer*. *Ra Ximhai* 8(3): 103-110.
- Sheldon MR, Fillyaw MJ, Thompson WD. 1996. The use and interpretation of the Friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International* 1(4): 221-228. <https://doi.org/10.1002/pri.66>
- Siegel S, Castellan NJ. 2005. Estadística no paramétrica aplicada a las ciencias de la conducta. Trillas. Ciudad de México, México.
- Sokal RR, Rohlf FJ. 2002. Introducción a la bioestadística. Reverté. Barcelona.
- Townend J. 2005. Practical statistics for environmental and biological scientists. Wiley. West Sussex, UK.
- Tukey JW. 1949. Comparing individual means in the analysis of variance. *Biometrics* 5: 99-114. <https://doi.org/10.2307/3001913>
- Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1(6): 80-83. <https://doi.org/10.2307/3001968>