

ERRORES FRECUENTES QUE SE PRESENTAN EN INVESTIGACIONES DESARROLLADAS EN LAS CIENCIAS AGRARIAS DEBIDO AL USO INCORRECTO DE MÉTODOS ESTADÍSTICOS

Frequently occurring errors in agricultural research due to the incorrect use of statistical methods

✉ Mario Varela Nualles^{1*}, ✉ Lucía Fernández Chuairey², ✉ José Antonio Pino Roque², ✉ Ileana Miranda Cabrera³

¹Instituto Nacional de Ciencias Agrícolas (INCA), San José de la Lajas, Mayabeque, Cuba

²Universidad Agraria de La Habana (UNAH), San José de la Lajas, Mayabeque, Cuba

³Centro Nacional de Sanidad Agropecuaria (CENSA), Apartado 10. San José de las Lajas, Mayabeque, Cuba

RESUMEN : Los objetivos del presente artículo fueron I) alertar sobre los errores estadísticos que se repiten, frecuentemente, en diferentes artículos científicos dentro de las Ciencias Agrarias, asociados a aspectos conceptuales y al uso incorrecto de las herramientas computacionales; específicamente en el contexto del Análisis de Varianza (ANOVA) y II) ofrecer metodologías de trabajo que permitan el adecuado uso de estos procedimientos estadísticos. Se abordan cuatro tipos de errores, el primero asociado al supuesto de normalidad, el segundo se vincula con pruebas de comparación múltiples no paramétricas, el tercero consiste en no considerar un error estándar único al informar los intervalos de confianza en un contexto ANOVA y el cuarto se relaciona con la interpretación del p-valor, al estar muy próximo al nivel de significación. Se incluye un ejemplo y se muestra la vía de cómo proceder mediante el uso de un software estadístico. En general, se plasman criterios y metodologías de trabajo que favorecen el correcto uso de estos procedimientos. Tener en consideración los aspectos abordados en este trabajo posibilitará evitar emitir conclusiones sesgadas, lo que contribuirá a obtener información significativa y valiosa de los resultados en los procesos de investigación y su comunicación a través de artículos científicos.

Palabras claves: análisis de varianza, errores estadísticos, sanidad vegetal, pruebas no paramétricas.

ABSTRACT : The objectives of this article were I) to warn about recurrent statistical errors that are frequently found in different scientific papers in agricultural research and associated with conceptual aspects and the misuse of computational tools, specifically, in the context of the analysis of variance (ANOVA) and II) provide work methodologies that allow the correct use of these statistical procedures. Four types of errors are approached: an error associated with the assumption of normality; a second one linked to non-parametric multiple comparison tests; another error consists of not considering a single standard error when reporting confidence intervals in an ANOVA context; and the fourth error is related to the interpretation of the p-value when it is very close to the level of significance. An example is included and shows how to proceed by using statistical software. In general, criteria and work methodologies are established that favor the correct use of these procedures. Taking into consideration the aspects approached in this work will make it possible to avoid issuing biased conclusions, which will contribute to obtaining significant and valuable information from the results in research processes and scientific papers.

Keywords: analysis of variance, statistical errors, plant health, non-parametric tests.

INTRODUCCIÓN

Por lo general, cuando se utiliza algún procedimiento estadístico en el análisis de datos, no siempre se posee asesoría de un especialista. En gran parte de los casos, se repite lo que otros hicieron sin verificar si el procedimiento que se aplicó fue apropiado, de ahí que, es muy común que un error estadístico pueda generalizarse.

En las Investigaciones Agropecuarias, uno de los Métodos Estadísticos más utilizados es el Análisis de Varianza (ANOVA), que como se conoce, requiere del cumplimiento de supuestos teóricos de normalidad de los datos, independencia de los errores, aleatorización y homogeneidad de varianzas (1). Uno de los errores más frecuentes en los que se incurre es no considerar el supuesto de normalidad dentro de cada grupo (o tratamiento), lo que conlleva a errores conceptuales y, por consiguiente, a tomar decisiones erróneas (2).

*Correspondencia a: Mario Varela Nualles. E-mail: varela@inca.edu.cu

Recibido: 09/10/2024

Aceptado: 11/12/2024

Declaración de conflicto de intereses: los autores declaran no poseer conflicto de intereses.

Contribución de los autores: Mario Varela Nualles: **Conceptualización, Análisis Formal, Investigación, Metodología, Escritura, Visualización, Escritura - borrador original, Redacción: revisión y edición.** Lucía Fernández Chuairey: **Conceptualización, Análisis Formal, Investigación, Metodología, Escritura, Visualización.** José Antonio Pino Roque: **Redacción: revisión y edición.** Ileana Miranda Cabrera: **Análisis Formal, Metodología y Redacción: revisión y edición.**



Este y otros errores fueron observados por especialistas biométricos del complejo científico-docente formado por la Universidad Agraria de La Habana (UNAH), el Centro Nacional de Sanidad Agropecuaria (CENSA), el Instituto Nacional de Ciencia Agrícola (INCA) y el Instituto de Ciencia Animal (ICA), al revisar trabajos donde se comparan tratamientos.

Los objetivos del presente artículo fueron I) alertar sobre los errores estadísticos que se repiten, frecuentemente, en diferentes artículos científicos dentro de las Ciencias Agrarias, asociados a aspectos conceptuales y mal uso de las herramientas computacionales; específicamente en el contexto del Análisis de Varianza (ANOVA) y II) ofrecer metodologías de trabajo, que permitan el correcto uso de estos procedimientos estadísticos.

PARTE ESPECIAL

Por ser el ANOVA una de las técnicas estadísticas mayormente empleada en las investigaciones agrícolas y considerando que un empleo incorrecto de sus supuestos teóricos, puede atentar contra el resultado de investigación, se enuncian un grupo de errores que se cometen y se dan orientaciones y vías para su correcto uso.

Error que se comete al no considerar que el supuesto de normalidad en el ANOVA es para la variable dentro de cada tratamiento o para los residuos del modelo

Con frecuencia, cuando se comprueba el cumplimiento de normalidad en el ANOVA, se realiza sin tener en consideración el tratamiento, o sea se prueba la normalidad de la variable respuesta y no se verifica el cumplimiento para cada tratamiento o grupo por separado. Al aplicar la prueba de Kolmogorov-Smirnov o Shapiro-Wilk a la totalidad de los datos sin tener en cuenta el tratamiento, lo más probable es que se rechace la hipótesis nula de normalidad, ya que en los datos está implícito el efecto del tratamiento.

Es conveniente resaltar que el procedimiento ANOVA, con factores fijos y tamaños de las muestras iguales funciona bastante bien, incluso cuando se viola el supuesto de normalidad, a menos que una o más de las distribuciones sean altamente asimétricas o las varianzas sean muy diferentes.

En el contexto ANOVA, lo correcto sería aplicar estas pruebas a cada tratamiento. Donde se busca no rechazar la hipótesis nula (H_0) (distribución es normal con $p\text{-valor} \geq 0,05$ u otro nivel de significación que se quiera probar). Regularmente, estas pruebas aparecen en los softwares estadísticos en análisis descriptivo, exploración de datos y pruebas de Normalidad. En caso de que todos los grupos distribuyan normalmente, se continúa con el procedimiento de ANOVA por técnicas paramétricas.

Sin embargo, ante el incumplimiento del supuesto de Normalidad en, al menos, uno de los grupos se recurre, en la mayor parte de los casos, a aplicar las técnicas no paramétricas, específicamente la Prueba de Kruskal-Wallis cuando se prueban varios tratamientos en presencia de un Diseño Completamente Aleatorizado. Este método se considera una generalización de la Prueba F de Fisher al caso no paramétrico (3, 4).

Algunos autores avalan que una alternativa muy utilizada es la transformación del dato, pero alegan que, al estar la variable en escala transformada, esto puede ocasionar problemas a la hora de interpretar el resultado, de ahí que se sugieran aplicar técnicas no paramétricas (5, 6, 7). Otros autores plantearon que se pueden transformar los datos y volver a probar los supuestos a los datos transformados y luego colocar en las tablas tanto el dato original como el transformado, para facilitar la interpretación, ya que siempre que sea posible, es conveniente emplear un método paramétrico (8).

Otro aspecto a considerar es que, en diversas investigaciones, el número de repeticiones por tratamiento es insuficiente, por lo que aplicar el test de normalidad, en cada tratamiento, puede conducir a resultados erróneos. Por otra parte, un tamaño muestral pequeño solo puede detectar efectos muy significativos y es también más susceptibles de no encontrar el efecto real que está presente en los datos (error de tipo II) (9). Adicionalmente, la distribución de una muestra pequeña tiende a desviarse de una distribución normal y el limitado tamaño hace, a menudo, imposible probar con rigor el supuesto de normalidad (6). Se sugiere aplicar el test de normalidad a los residuales del modelo y en caso de incumplimiento del supuesto, proceder al uso de una técnica no paramétrica.

Se muestra un ejemplo hipotético, en el que se evalúa el efecto de tres tratamientos sobre la altura de plantas de plátano (*Musa spp.*) cultivar 'Gran Enano' (T1. Plantas con micorrizas, T2 Testigo plantas sin micorrizas, T3 Plantas con micorrizas + Biochar). Para ello, se consideran 10 repeticiones por tratamiento. Para realizar el análisis del cumplimiento de normalidad, la base de datos se confecciona con los tratamientos en columna, en este caso se utilizó como software estadístico el SPSS (10). (Fig. 1)

Seguir este procedimiento en el SPSS (Analizar, Estadísticos Descriptivos, Explorar, Colocar las variables a las que se les desea analizar normalidad en lista de variables dependientes, seleccionar en la pestaña gráficos los gráficos con prueba de normalidad), permite obtener la significación en cada uno de estos grupos y, en dependencia de este resultado obtenido, proceder. Si resulta que todos los grupos tienen distribución normal se continúa con técnicas paramétricas (ANOVA). En el ejemplo que se muestra, se incumple el supuesto de normalidad para el tratamiento 3 ($p=0,017$), por lo que se sugiere aplicar la prueba no paramétrica de Kruskal-Wallis (Tabla 1).

En el software SPSS, para realizar la prueba Kruskal-Wallis, los datos deben colocarse ubicando un tratamiento debajo del otro, una columna referirá a la variable dependiente (en medida de escala) y la otra denominada factor, debe indicar el tratamiento (declarada en medida Nominal) (Fig. 2).

Al realizar la prueba Kurkall-Wallis en el ejemplo se obtuvo $p\text{-valor} < 0,05$, lo cual indica que no se acepta H_0 , es decir, se encontraron diferencias estadísticamente significativas entre los grupos (Tabla 2). Sin embargo, ¿cuáles grupos difieren? Es necesario aplicar una prueba de Comparación Múltiple no paramétrica.

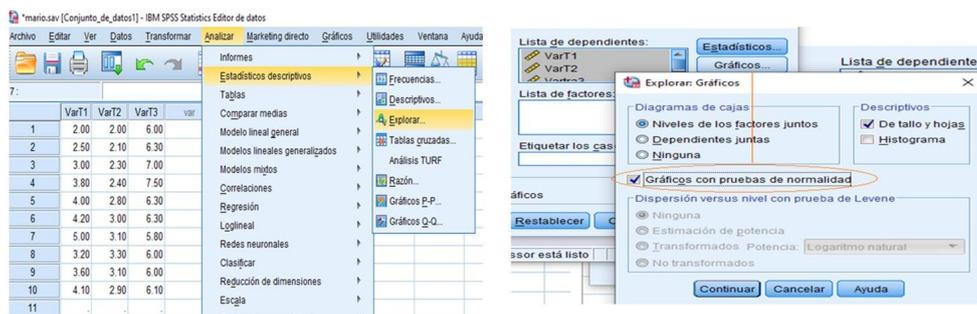


Figura 1. Gráfico para aplicar la prueba de Kolmogorov-Smirnov o Shapiro-Wilk, en el SPSS. / Graph to apply Kolmogorov-Smirnov or Shapiro-Wilk test in SPSS

Tabla 1. Prueba de Normalidad por tratamiento / Normality test for each treatment

| Tratamiento | Kolmogorov-Smirnov | | | Shapiro-Wilk | | |
|-------------|--------------------|----|--------|--------------|----|-------|
| | Estadígrafo | gl | p | Estadígrafo | Gl | P |
| 1 | 0,128 | 10 | 0,200* | 0,979 | 10 | 0,960 |
| 2 | 0,186 | 10 | 0,200* | 0,913 | 10 | 0,302 |
| 3 | 0,323 | 10 | 0,004 | 0,805 | 10 | 0,017 |

* Límite inferior de la significación verdadera/ Lower limit of true significance



Figura 2. Declaración de variables para realizar prueba Kruskal-Wallis/ Declaration of variables to perform Kruskal-Wallis test

Tabla 2. Resultado de la prueba no paramétrica Kruskal-Wallis para muestras independientes / Result of the Krukall-Wallis nonparametric test for independent samples

| Resumen de contraste de hipótesis | | | |
|--|---|---------------|---------------------------|
| Hipótesis nula | Prueba | Significación | Decisión |
| La distribución de variable es la misma entre las categorías de factor | Kruskal-Wallis para muestras independientes | 0,0001 | Rechace la hipótesis nula |

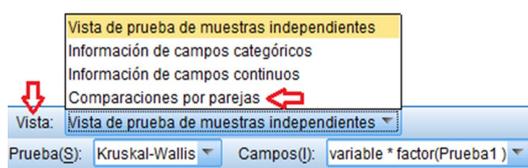


Figura 3. Vista del SPSS para comprobar diferencias entre tratamientos después de efectuar la Prueba no paramétrica de k muestras independientes / View of SPSS to check differences between treatments after performing the non-parametric test of k independent samples

Error que se comete al aplicar la Prueba de Mann-Witney para conocer los pares de tratamientos que difieren, luego de encontrar diferencias significativas mediante la prueba de Kruskal-Wallis

Al encontrar diferencias entre tratamientos, es preciso conocer cuáles tratamientos difieren entre sí, erróneamente se aplica la prueba no paramétrica de Mann-Witney, que está diseñada para comparar solamente dos tratamientos (no múltiples tratamientos) (11), de ahí que es preciso, en tal caso, aplicar una prueba de Comparación Múltiple, para técnicas no paramétricas.

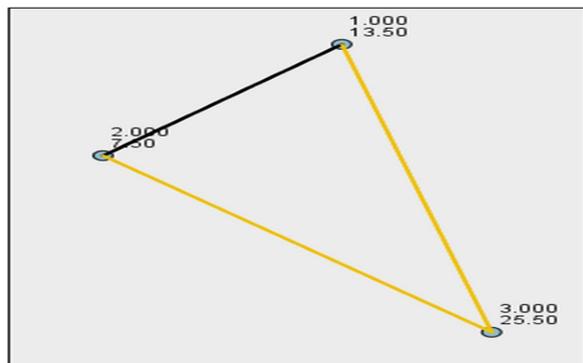
En el caso del SPSS, una vez procesada la prueba no paramétrica de k muestras independientes, al dar doble clic

en la tabla de resultados (Tabla 2) y automáticamente se activa una nueva tabla, y marcamos comparación por parejas (Fig. 3).

La salida es la comparación por pares de los tratamientos y la figura correspondiente (Tabla 3, Fig. 4). En el grafico se informan los rangos promedios de cada tratamiento, así como los pares de tratamientos para los cuales resulta significativa la diferencia de rangos entre tratamientos (Estadístico). No se encontraron diferencias entre los tratamientos 1 y 2. Contrariamente, se encontraron diferencias estadísticas entre los tratamientos 2 y 3 y entre los tratamientos 1 y 3.

Tabla 3. Resultado de la prueba de Comparación Múltiple no Paramétrica. / Nonparametric Multiple Comparison test result

| Muestra 1-Muestra 2 | Estadístico | ES | DS | p | p-ajustado |
|---------------------|-------------|-------|--------|-------|------------|
| T2-T1 | 6 | 3,932 | 1,526 | 0,127 | 0,381 |
| T2-T3 | -18 | 3,932 | -4,578 | 0,000 | 0,000 |
| T1-T3 | -12 | 3,932 | -3,052 | 0,002 | 0,007 |

**Figura 4.** Rango promedio por tratamiento / Average range per treatment

Otro error muy común, está en la interpretación o vocabulario que se emplea cuando en una prueba de hipótesis, no se rechaza H_0 ($p\text{-valor} > 0,05$) y con ello se infiere que los grupos o tratamientos son iguales. En tal caso, lo correcto es decir que no se encontraron evidencias suficientes para concluir que los grupos son diferentes. No encontrar diferencias puede estar asociado a otros factores, por ejemplo, gran variabilidad dentro de tratamientos, o sea, error experimental.

En tal sentido, algunos autores plantearon que hay dos opciones para tomar una decisión, y estas son: "rechazar H_0 ", si la información de la muestra favorece la hipótesis alternativa o "no rechazar H_0 " o "negarse a rechazar H_0 ", si la información de la muestra es insuficiente para rechazar la hipótesis nula (12).

Error que se comete al presentar simultáneamente los resultados del ANOVA y los intervalos de confianza con un error estándar diferente por tratamiento

Cuando se realiza un ANOVA se hace bajo el supuesto de Homogeneidad de Varianzas, por tanto, si se necesita informar los intervalos de confianza para cada tratamiento, lo correcto es considerar un único error estándar (el asociado al ANOVA). Los intervalos de confianza deben tener la misma amplitud y se debe considerar como error estándar el siguiente:

$$E.E = \sqrt{\frac{CME}{n}}$$

Donde CME representa el Cuadrado Medio del Error asociado al ANOVA y n el número de repeticiones por tratamiento para experimentos balanceados.

Incorrecta interpretación del valor de p

Resulta igualmente conveniente informar en una publicación el valor exacto de p , para evitar diferente

interpretación, sobre todo cuando al fijar como zona de rechazo un valor de $p < 0,05$ y el p -valor encontrado está próximo a este valor, como por ejemplo cuando alcanza valor de 0,049 o de 0,051.

En tal sentido, diversos autores refirieron que un valor p no significativo puede representar que un resultado es realmente nulo, que es un efecto que no tiene poder estadístico suficiente para su evaluación o que es un efecto ambiguo (9, 13). Para interpretar un resultado no significativo, como una evidencia en contra de la hipótesis, se necesitaría demostrar que esa evidencia es significativa. Esto supone que resultados que se encuentren cercanos al 0,05 no deban asumirse como no satisfactorios, cuando realmente proporcionan evidencia preliminar que requiere atención adicional. Es muy importante en estos casos tener en consideración la experiencia del investigador e incluso pudiera aumentarse el valor de p para detectar más diferencias.

CONCLUSIONES

El uso incorrecto de procedimientos estadísticos conduce a interpretaciones erróneas, siendo de vital importancia tener en consideración los aspectos abordados en este trabajo para evitar obtener conclusiones sesgadas y por ende, lograr una información significativa y valiosa de los resultados en los procesos de investigación y artículos científicos.

AGRADECIMIENTOS

La investigación que da origen a los resultados presentados en este artículo recibió fondos de la Oficina de Gestión de Fondos y Proyectos Internacionales bajo el código PN223LH010-066 correspondiente al proyecto "Desarrollo de Modelos Estadísticos-Matemáticos y herramientas de la Agricultura Inteligente, para su aplicación en la descripción y gestión de Procesos Agrarios" del Programa Nacional de Ciencias Básicas y Naturales.

REFERENCIAS

1. Dixon W J, Massey FJ. Introducción al análisis estadístico. Editorial Pueblo y Educación (Cuba). Segunda edición. 1974. 489 pp.
2. Goss-Sampson M A. Análisis Estadístico con JASP: Una guía para estudiantes. Editorial FUOC. Primera versión traducida. 2019. 128 pp.
3. Matamoros-Pinel RA, Caballe Ros-Marque A. Errores conceptuales más comunes en publicaciones científicas. Revista Medicina Veterinaria y Zootecnia 2017;12 (3): 1211-1290. DOI: <https://doi.org/10.21615/cesmvz.12.3.4>
4. Ortega-Páez E, Ochoa-Sangrador C, Molina-Arias M. Pruebas no paramétricas. Evid. Pediatr. 2021; 17(37):1-7.

5. Pérez-Pelea I. ¿Cómo proceder ante el incumplimiento de las premisas de los métodos paramétricos? o ¿cómo trabajar con variables biológicas no normales? *Revista del Jardín Botánico Nacional* 2018 40(1): 1-12.
6. De Calzadilla J, Guerra W, Torres V. El uso y abuso de transformaciones matemáticas. Aplicaciones en Análisis de Varianza. *Revista Cubana de Ciencia Agrícola*. 2002; 36(2):103-106.
7. Gómez-Gómez M, Danglot - Banck C, Vega-Franco I. Sinopsis de pruebas estadísticas no paramétricas. Cuándo usarlas. *Revista Mexicana de Pediatría*. 2003; 70(2): 91-99.
8. Miranda I. Estadística Aplicada a la Sanidad Vegetal. Editorial EDICENSA. Centro Nacional de Sanidad Agropecuaria (CENSA), Cuba. 2011. 176 pp.
9. Campo-Albendea L, Muriel-García A. Diez errores estadísticos frecuentes que tener en cuenta al escribir o revisar un artículo. *Revista Enfermería Intensiva*. 2021; 32(1):42-44. <https://doi.org/10.1016/j.enfi.2020.09.002>
10. IBM SPSS: Statistics. Versión 22. 2014. <https://ibm-spss-statistics-base.mrdownload.com/es/para-windows/app/descargar/>
11. Molina M. Prueba U de Mann-Whitney. *Ciencias o Letras. Rev. Elect. Anestesiología*. 2023; 15 (5):1-6. <https://dialnet.unirioja.es/descarga/articulo/8993603.pdf>
12. Illowsky B, Dean S. Introducción a la estadística. Editorial OpenStax, Houston, Texas. 2022. 260 pp.
13. Mendivelso FO, Rodríguez I M. Definición e interpretación adecuada del p-valor e intervalos de confianza en investigación clínica. *Revista Médica Sanitas*. 2018; 21(4):193-196. <http://dx.doi.org/10.26852/01234250.28>