

Referencia

López Bautista, E. A. y González Ramírez, B. H. (2024). Análisis de experimentos factoriales para la investigación agronómica mediante Infostat. *Revista Científica del Sistema de Estudios de Postgrado*. 7(2). 17-30.

DOI: <https://doi.org/10.36958/sep.v7i2.299>

Análisis de experimentos factoriales para la investigación agronómica mediante Infostat

Analysis of factorial experiments for agronomical reseach using Infostat

Ezequiel Abraham López Bautista

Profesor titular de la Facultad de Agronomía
Universidad de San Carlos de Guatemala
ealbautis@gmail.com

<https://orcid.org/0000-0002-4724-4251> 

Byron Humberto González Ramírez

Profesor titular y Director del Centro de Telemática
(CETE) de la Facultad de Agronomía
Universidad de San Carlos de Guatemala
byrong.gt@gmail.com

<https://orcid.org/0000-0002-2787-9508> 

Recibido: 05/02/2024

Aceptado: 26/09/2024

Publicado: 26/11/2024

Resumen

OBJETIVO: mostrar a los investigadores el análisis estadístico de un experimento bifactorial empleando software. **MÉTODO:** en la descripción, se emplean datos de rendimiento de avena, cultivada en tres fechas de siembra y cuatro niveles de nitrógeno, en un diseño de bloques completos al azar y arreglo combinatorio. Se realizó el análisis de la varianza (ANOVA), verificación de los supuestos, gráfico de las interacciones y análisis post ANOVA; empleando el software Infostat v.2020®. Para la verificación de los supuestos del modelo estadístico fue empleada la prueba de Shapiro-Wilk (para verificar la normalidad) y el gráfico de dispersión entre valores predichos y los residuos estudentizados (para verificar la homocedasticidad y la independencia). Como prueba post ANOVA se empleó la comparación de medias DGC. **RESULTADOS:** la tabla resumen del ANOVA para un experimento bifactorial fue generada en el Infostat v.2020® junto con el valor del coeficiente de variación (CV) como indicador de precisión del experimento. Los factores principales y la interacción fueron significativos. Los p valores asociados a las fechas, niveles de nitrógeno, e interacción fueron de <0.0001, 0.0034 y 0.0325 respectivamente. El CV fue de 24.67%, con intervalo de confianza, generado por simulación de

Las opiniones expresadas en el artículo son responsabilidad exclusiva de los autores y no necesariamente representan la posición oficial de la USAC y sus miembros. La obra está protegida por la Ley de Derechos de Autor y Derechos Conexos emitida en el decreto No. 33-98 por el Congreso de la República de Guatemala.

<https://revistasep.usac.edu.gt>

22.27-27.07%. En la revisión de supuestos, estos fueron cumplidos. De acuerdo con la prueba DGC, los mejores rendimientos fueron obtenidos en la primera fecha de siembra y, cuando se aplicó nitrógeno, generándose un modelo cuadrático. **CONCLUSIÓN:** El uso de *software* permite realizar el análisis de manera fácil, rápida y confiable. También es posible disponer de más tiempo para la interpretación de los resultados del ANOVA.

Palabras clave

experimentos factoriales, *software* estadístico, análisis de la varianza, prueba de comparación de medias, estadística experimental

Abstract

OBJECTIVE: to show researchers the statistical analysis of a two-factor experiment using software. **METHOD:** In the description, oat yield data are used, grown on three sowing dates and four nitrogen levels, in a randomized complete block design and combinatorial arrangement. The analysis of variance (ANOVA), verification of assumptions, interaction graph and post-ANOVA analysis were performed; using the Infostat v.2020® software. To verify the assumptions of the statistical model, the Shapiro-Wilk test was used (to verify normality) and the scatter graph between predicted values and studentized residuals (to verify homoscedasticity and independence). The DGC comparison of means was used as a post-ANOVA test. **RESULTS:** The summary table of the ANOVA for a two-factor experiment was generated in Infostat v.2020® along with the coefficient of variation (CV) value as an indicator of the precision of the experiment. The main factors and the interaction were significant. The p values associated with the dates, nitrogen levels, and interaction were <0.0001, 0.0034, and 0.0325. The CV was 24.67%, with a confidence interval generated by simulation of 22.27-27.07%. In the review of assumptions, these were met. According to the DGC test, the best yields were obtained on the first sowing date, and when nitrogen was applied, generating a quadratic model. **CONCLUSION:** The use of software allows for easy, fast, and reliable analysis. It is also possible to have more time for the interpretation of the ANOVA results.

Keywords

factorial experiment, statistical software, analysis of variance, mean difference test, polynomial regression

Introducción

Los experimentos factoriales poseen dos o más factores, lo más común es emplear dos o tres como máximo, cada uno con distintos niveles posibles, debido a la complejidad de interpretación de las interacciones (Martínez, 1988). Todas las factibles combinaciones de los niveles de los factores (considerados de efectos cruzados) se ubican en las unidades experimentales.

Una investigación de este tipo permite al investigador estudiar el efecto principal de cada factor, así como los efectos de las interacciones entre ellos sobre la variable dependiente o de respuesta. Es posible clasificar estos ensayos como simétricos, como el caso de un experimento factorial 3², donde se tienen dos factores cada uno con tres niveles. El experimento factorial tendría nueve combinaciones o tratamientos en total. También los experimentos se clasifican como asimétricos, como el caso de un factorial 3 x 4 (Martínez, 1988). En ambos casos, las interacciones entre los factores son regularmente las más significativas y representan el objetivo central de los investigadores.

Los experimentos factoriales fueron utilizados en el siglo XIX por John Bennet Lawes y Joseph Henry Gilbert de la Estación Experimental Rothamsted, Londres, Inglaterra (Yates & Mather, 1963). Fisher (1926) argumentó que los experimentos «complejos» como los ensayos factoriales eran más eficientes que estudiar un factor a la vez.

Oehlert (2000) y Montgomery (2013) citaron que dentro de las ventajas de los experimentos factoriales en comparación con los simples, se tiene el ahorro de recursos y lo más importante, el análisis de las interacciones, como una mejor aproximación a lo que sucede en la naturaleza. Como inconvenientes, se tiene que al incrementar el número de niveles y/o factores, crece demasiado el número de interacciones, con algunas de poco interés práctico. Además, se demanda de mayor área experimental, que ocasiona problemas de manejo del ensayo.

El desarrollo de *software* en esta era de globalización, ha acentuado nuevos modos de transformación del conocimiento en la educación y el campo laboral (Parmar et al., 2022). Las herramientas computacionales, principalmente las de acceso libre o construidas con base en éstas para fines educativos, han abierto oportunidades para los centros de enseñanza e investigación en países en desarrollo, en términos de proporcionar acceso al análisis de datos confiable y a bajo costo.

En este contexto, el uso de *software* como Infostat® (Di Rienzo et al., 2020), ampliamente utilizado en América Latina para el análisis de datos experimentales, y también empleado en la Facultad de Agronomía de la Universidad de San Carlos de Guatemala (FAUSAC) desde el año 2002, ayuda al usuario final a inferir los resultados de una manera simple y ordenada para tomar decisiones rápidas. Así, el docente dispondrá de más tiempo para discutir la interpretación de los resultados. Este escrito será de utilidad para que estudiantes e investigadores conozcan el análisis de experimentos factoriales para la investigación agronómica utilizando *software* estadístico.

Objetivo

Mostrar a los docentes, estudiantes e investigadores del área agronómica, el análisis estadístico de un experimento bifactorial dispuesto en un diseño de bloques completos al azar y arreglo combinatorio, empleando Infostat® v. 2020

Materiales y métodos

Infostat® es un *software* distribuido bajo licencia, que también dispone de una versión estudiantil, sin costo alguno. Fue creado bajo el soporte del entorno de programación R, con fines educativos. Esto hace que su uso sea inmediato en la versión libre y su costo sea accesible cuando se adquiere mediante licencia por estudiantes, docentes o investigadores. Para realizar el análisis de datos, el usuario necesita importar un conjunto de datos ordenado y depurado en una hoja de cálculo electrónica de *Excel*.

Para ilustrar el uso de Infostat® fueron empleados los datos citados por Parmar et al. (2022), referentes a un experimento bifactorial sobre rendimiento de granos de avena (*Avena sativa*), con arreglo combinatorio dispuesto en un diseño de bloques completos al azar (con seis repeticiones). Los factores evaluados fueron fechas de siembra (FS1, FS2 y FS3) y cuatro niveles de nitrógeno (NO, N1, N2 y N3), para un total de 12 combinaciones (tratamientos). El conjunto de datos obtenido al realizar el ensayo, referente a la producción de grano de avena (expresada en arobas por unidad experimental) se presenta en la tabla 1:

Tabla 1

Distribución de los rendimientos de granos de avena por tratamiento, fecha de siembra, niveles de nitrógeno y bloques.

Tratamiento	FS	N	Bloques					
			I	II	III	IV	V	VI
T1	FS1	N0	8.3	10.3	8.0	8.0	6.0	8.5
T2	FS1	N1	9.3	9.0	9.5	11.7	11.3	10.7
T3	FS1	N2	11.3	11.5	11.3	11.7	14.7	15.0
T4	FS1	N3	10.5	15.7	10.5	11.7	10.3	15.0
T5	FS2	N0	5.7	4.5	8.3	8.5	5.3	4.3
T6	FS2	N1	4.7	5.3	8.0	9.3	5.3	6.7
T7	FS2	N2	5.3	5.5	8.0	8.7	8.5	11.3
T8	FS2	N3	6.5	8.3	8.0	9.0	6.3	8.5
T9	FS3	N0	5.0	4.7	1.5	3.0	3.7	4.5
T10	FS3	N1	7.0	8.3	2.5	3.5	6.7	4.5
T11	FS3	N2	3.3	3.3	2.5	1.0	6.3	8.0
T12	FS3	N3	2.7	4.3	1.3	2.5	4.0	7.7

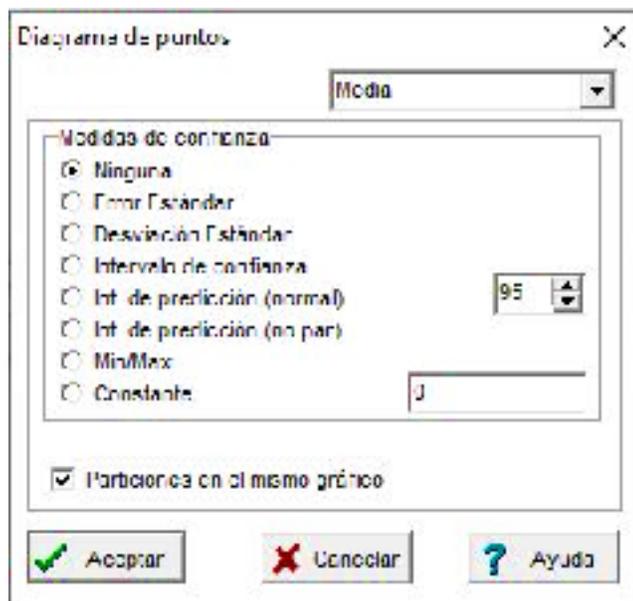
Nota: tomado de Parmar et al. (2022)

El procedimiento para analizar los datos en Infostat® se detalla a continuación:

- Preparar el conjunto de datos para su exportación desde *Excel* a Infostat. Los datos están disponibles en este enlace <http://cete.fausac.gt/wp-content/uploads/2024/09/factoriales.xlsx>
- Importar los datos a Infostat, al ubicar en el menú principal, la opción «edición» y luego presionar en: «Pegar con nombres de columnas» (brindar atención a copiar desde *Excel* el nombre de las columnas y los datos).
- Construir el gráfico de las interacciones, mediante el uso del gráfico de puntos, ubicado en el menú «gráficos». En «Criterios de clasificación (optativa)» coloque los niveles del factor nitrógeno (N), en variables a graficar VR y en «Particiones» las fechas de siembra (FS). Aparecerá un recuadro como en la figura 1 (es necesario marcar en «medidas de confianza»: Ninguna y activar «Particiones en el mismo gráfico» (Balzarini et al., 2008).

Figura 1

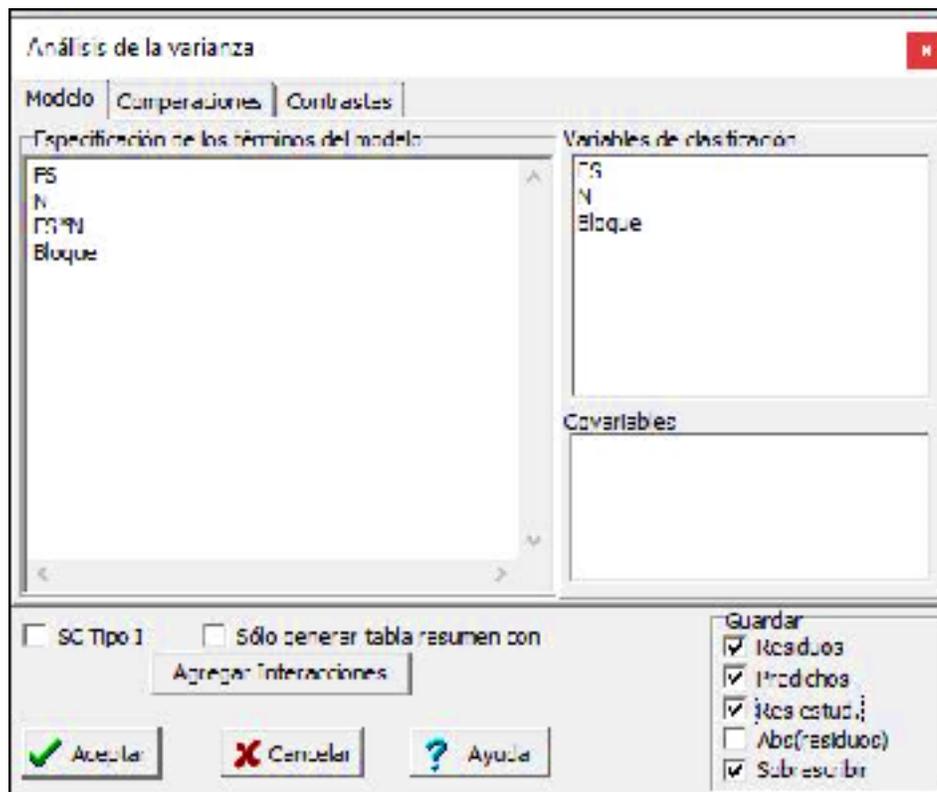
Diagrama de puntos



- Posteriormente se realiza el ANOVA, con un nivel de significancia del 5%. En el menú «Estadísticas» seleccionar «Análisis de la Varianza» y especificar los términos del modelo tal como se muestra en la figura 2:

Figura 2

Especificación de los términos del modelo estadístico-matemático



Agregar la interacción FS*N y marcar los residuos, predichos y residuos estudentizados, que servirán para realizar el análisis de los residuos del modelo estadístico matemático para verificar su cumplimiento.

El modelo estadístico-matemático utilizado es el siguiente (Montgomery, 2013):

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \varepsilon_{ijk}$$

Con:

i = 1, 2, 3 fechas de siembra

j = 1,2,3,4 niveles de nitrógeno

k = 1,2,3,4,5,6 bloques

Siendo:

Y_{ijk} =rendimiento de granos de avena medidos en la ijk -ésima unidad experimental

μ =media general del rendimiento de grano de avena

α_i =efecto de la i -ésima fecha de siembra de la avena

β_j =efecto del j -ésimo nivel de nitrógeno

$(\alpha\beta)_{ij}$ =efecto de la interacción entre la i -ésima fecha de siembra y el j -ésimo nivel de nitrógeno

γ_k =efecto del k -ésimo bloque o repetición

ε_{ijk} =error experimental asociado a la ijk -ésima unidad experimental

- Para la verificación de los supuestos del modelo fueron aplicados los siguientes procedimientos.

Normalidad de los residuos: fue evaluada la hipótesis nula H_0 : siguen una distribución normal con media 0 y varianza 1. Para ello, fue empleada la prueba de Shapiro-Wilk modificada (Di Rienzo et al., 2020).

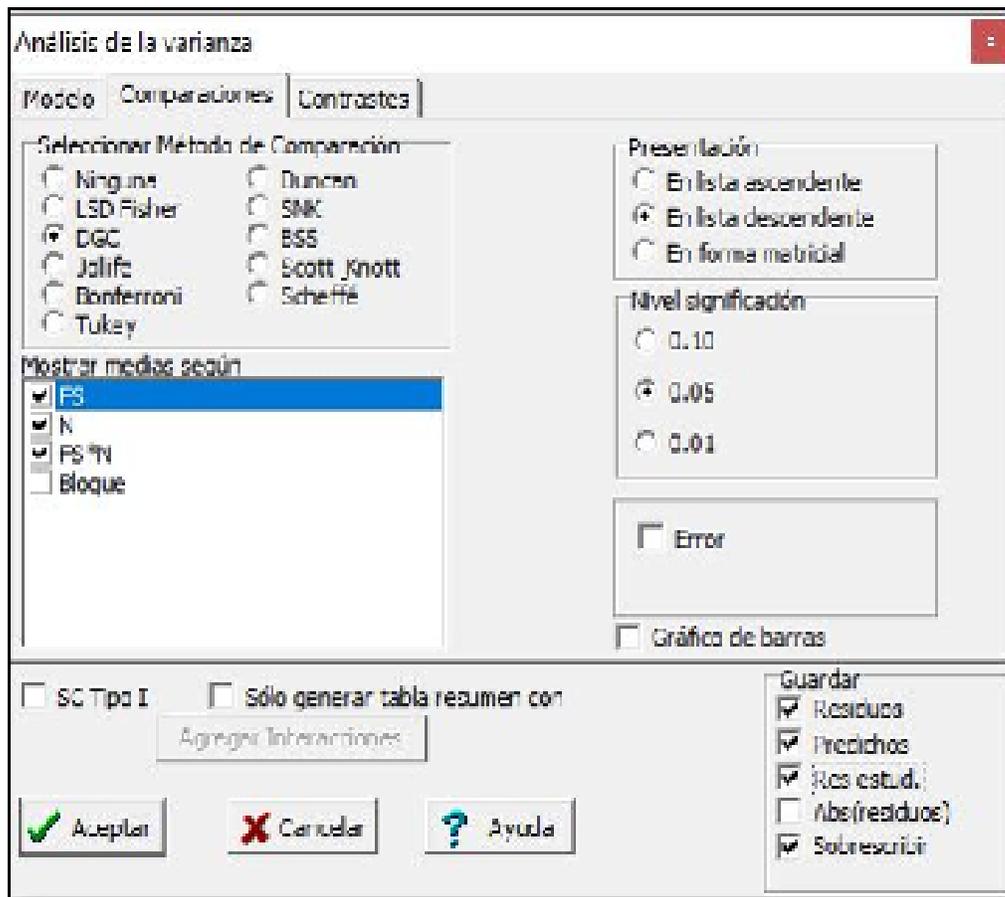
Homocedasticidad (igualdad de varianzas) e independencia de los residuos. En este caso fue construido un gráfico de dispersión, considerando en el eje X a los valores predichos por el modelo y en el eje Y a los residuos estudentizados (RE).

Los valores de estos últimos están comprendidos entre los límites de -3 y 3, que de acuerdo con el teorema de Tchebyshev, concentran aproximadamente el 99% de los datos. Fuera de estos límites, cualquier par graficado se considera como atípico (outlier) o extremo.

- Como análisis posterior al ANOVA fue aplicada la prueba de comparación de medias de Di Rienzo, Guzmán y Casanoves identificada como DGC (Di Rienzo et al., 2002), que está basada en el análisis multivariado de conglomerados (Cluster analysis), tal como se muestra en la figura 3.

Figura 3

Especificación de la prueba de DGC en Infostat



Análisis de la varianza

Modelo | Comparaciones | **Contrastes**

Seleccionar Método de Comparación:

- Ninguna
- LSD Fisher
- DGC
- Jalife
- Bonferroni
- Tukey
- Duncan
- SNK
- BSS
- Scott_Knott
- Scheffé

Mostrar medias según:

- PS
- N
- PS*N
- Bloque

Presentación:

- En lista ascendente
- En lista descendente
- En forma matricial

Nivel significación:

- 0.10
- 0.05
- 0.01

Error

Gráficos de barras

SC Tipo I Sólo generar tabla resumen con

Agregar Interacciones

Guardar:

- Residuos
- Predichos
- Res. estud.
- Abs(residuos)
- Sol. res. stud.

Para el análisis de polinomios ortogonales fue ajustado un modelo polinomial de grado 2 (cuadrático), considerando como variable independiente los niveles crecientes de nitrógeno y como variable dependiente el rendimiento de grano de avena, para cada fecha de siembra.

Resultados y discusión

En la tabla 2 se presenta el resumen del ANOVA con un 5% de significancia, proporcionado por el Infostat, así mismo los valores de coeficiente de variación ($CV = 24.67\%$) y el coeficiente de determinación ($R^2 = 78\%$).

Tabla 2

Resumen del análisis de la varianza para la variable rendimiento de granos de avena

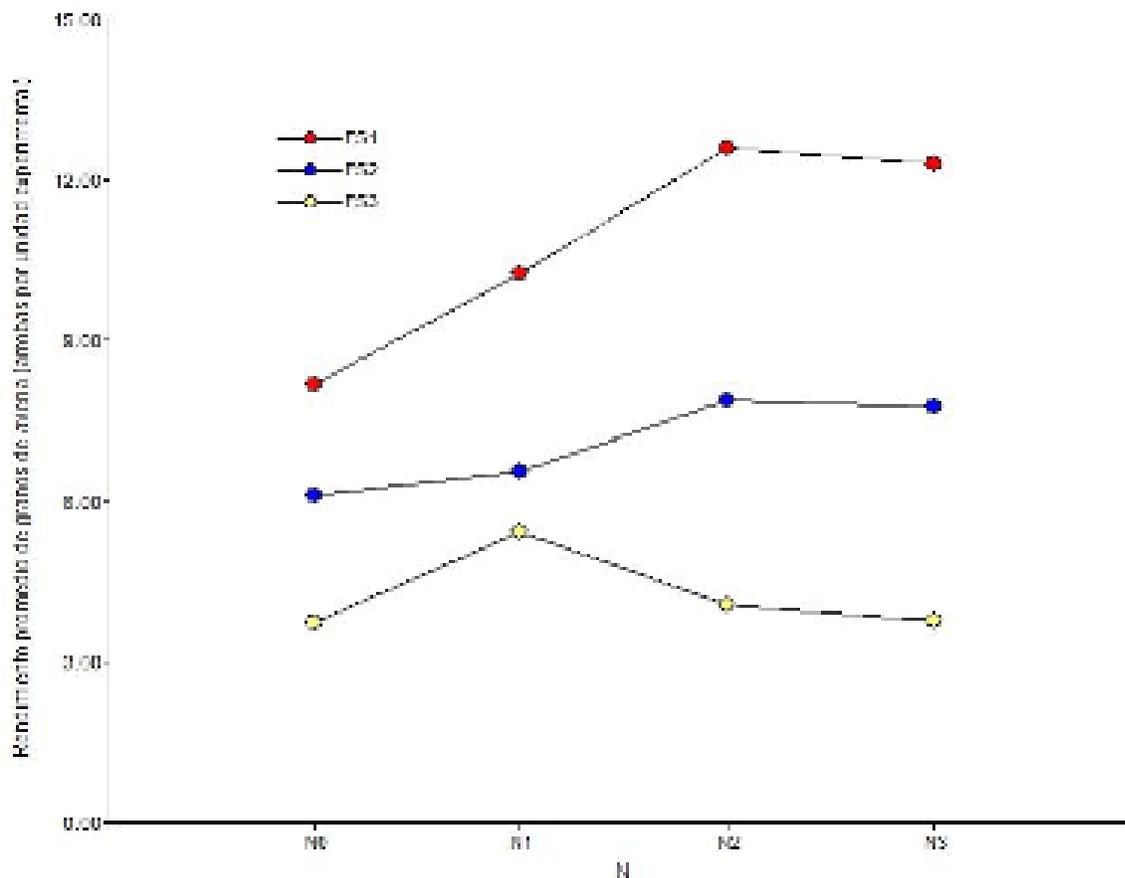
Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios	Valor de F	Valor de p
Fechas de siembra	523.44	2	261.72	78.97	<0.0001
Niveles de nitrógeno	50.98	3	16.99	5.13	0.0034
Interacción	49.80	6	8.30	2.50	0.0325
Bloque	35.77	5	7.15	2.16	0.0720
Residuo	182.29	55	3.31		
Total	842.29	71			

De acuerdo con el ANOVA, hubo efecto significativo (valores de $p < 0.05$) de los factores principales. Se identifica la fecha de siembra con la mayor significancia. Le siguen los niveles crecientes de nitrógeno. Además, hubo efecto de la interacción, que en los experimentos factoriales, es el principal interés. Respecto al $CV = 24.67\%$ está considerado como aceptable para experimentos agronómicos realizados en campo abierto. El valor de $R^2 = 78\%$ indica que el 78% de la variabilidad total del rendimiento de granos de avena se explica por la parte fija del modelo estadístico (factores, interacciones y bloques) y el 22% debido al azar.

Para analizar las interacciones de manera más detallada fue construida la figura 4. En ella, se confirma que el efecto de la interacción es significativo, debido a que las líneas no son paralelas. Además, se observa que el nivel fecha de siembra 1 presenta los valores más altos de rendimiento promedio de grano de avena. Se alcanza su máximo, cuando es aplicado el nivel 2 de nitrógeno. Por otra parte, con la fecha de siembra 1 se presentan los rendimientos promedio más bajos.

Figura 4

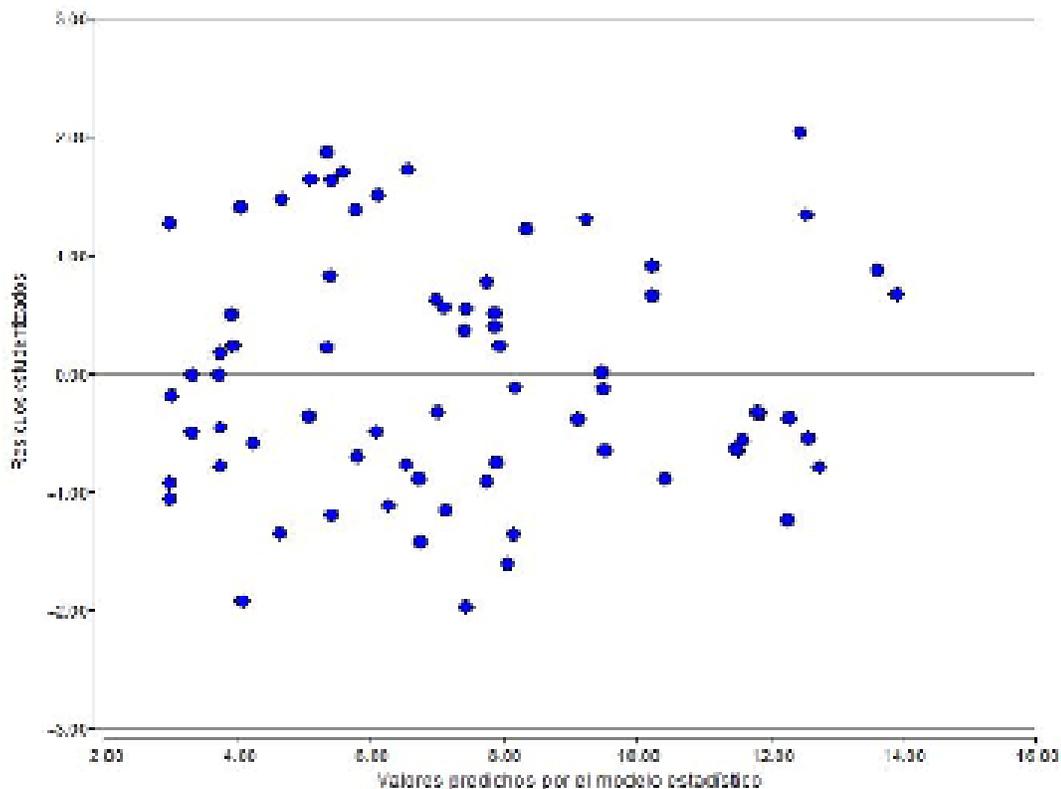
Gráfico de las interacciones entre la fecha de siembra y los niveles de nitrógeno



Respecto al cumplimiento de los supuestos del modelo estadístico del experimento, la prueba de Shapiro-Wilk proporcionó un valor de $p = 0.0539$. Se concluye que los residuos siguen la distribución normal.

En la figura 5 se presentan el gráfico de dispersión para verificar la homocedasticidad y la independencia de los residuos. Se observa que la nube de puntos presenta un comportamiento aleatorio (sin alguna tendencia definida). De esta forma, se concluye que las varianzas son homogéneas (se cumple el supuesto de homocedasticidad) y además son independientes (o sea, que no están correlacionados). Esto es importante, para darle validez a los resultados experimentales.

Figura 5
Gráfico de dispersión de residuos



Finalmente, se presenta la tabla 3 con el resumen de la prueba de comparación de medias de DGC, de donde se concluye que con la fecha 1 de siembra se tienen los mejores resultados de rendimiento promedio, con los niveles de nitrógeno N1, N2 y N3, superior al N0 (sin aplicación de nitrógeno).

Tabla 3
Resumen del análisis de la prueba de comparación de medias de DGC

Fecha de siembra	Nivel de nitrógeno	Media del rendimiento	Grupo DGC
FS1	N2	12.58	a
FS1	N3	12.28	a
FS1	N1	10.25	a
FS1	N0	8.18	b
FS2	N2	7,88	b
FS2	N3	7.77	b
FS2	N1	6.55	b
FS2	N0	6.10	b
FS3	N1	5.42	b
FS3	N2	4.07	c
FS3	N3	3.75	c
FS3	N0	3.73	c

Conclusión

El uso de Infostat v. 2020® permitió realizar los análisis estadísticos de datos experimentales de manera fácil, rápida y confiable. Derivado de ello, se dispone de más tiempo para la interpretación y discusión de los resultados con los usuarios. Además, hace posible la verificación de los supuestos del modelo y generar gráficas de alta calidad.

Referencias

Balzarini, M., González, L., Tablada, M., Casanoves, F., Di Rienzo, J., & Robledo, C. (2008). Infostat. Manual del Usuario. Editorial Brujas.

https://repositorio.catie.ac.cr/bitstream/handle/11554/10346/Manual_INFOSTAT_2008.pdf

Di Rienzo, J., Cassanoves, F., Balzarini, M., González, L., Tablada, M., & Robledo, C. (2020). Infostat (Versión 2020) [Software]. Centro de Transferencia InfoStat.

<http://www.infostat.com.ar>

Di Rienzo, J., Guzmán, A., & Casanoves, F. (2002). A Multiple-Comparisons Method Based on the Distribution of the Root Node Distance of a Binary Tree. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(2), 129–142. JSTOR. <https://www.jstor.org/stable/1400690>

Fisher, R. (1926). The arrangement of field experiments. 33, 503–515.
<https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15191/1/48.pdf>

Oehlert, G. (2000). A First Course in Design and Analysis of Experiments. W. H. Freeman and Company. <https://www.semanticscholar.org/paper/A-first-course-in-design-and-analysis-of-Oehlert/34703d23bd4283965894c3ce92d5ce5e1e397ac0>

Parmar, R. S., Kathiriya, D. R., & Kamani, G. J. (2022). Analysis of factorial experiments for agricultural research using digital tools. *Gujarat Journal of Extension Education*, 33(1), 150–154. <https://doi.org/10.56572/gjoe.2022.33.1.0030>

Yates, F., & Mather, K. (1963). Ronald Aylmer Fisher, 1890-1962. Biographical Memoirs of Fellows of the Royal Society, 9, 91–129. <https://doi.org/10.1098/rsbm.1963.0006>

Sobre los autores

Ezequiel Abraham López Bautista

Es Ingeniero Agrónomo. Doctor en Ciencias. Profesor titular de la Sub área de Métodos de Cuantificación e investigación de la Facultad de Agronomía, USAC.

Byron Humberto González Ramírez

Es Ingeniero Agrónomo. Ph. D., Profesor titular y Director del Centro de Telemática (CETE) de la Facultad de Agronomía, USAC.

Financiamiento de la investigación

Con recursos propios.

Declaración de intereses

Declaramos no tener conflicto alguno de intereses, que influyeran en los resultados obtenidos o las interpretaciones propuestas.

Declaración de consentimiento informado

El estudio se realizó respetando el Código de ética y buenas prácticas editoriales de publicación.

Derecho de uso

Copyright (c) (2024) Ezequiel Abraham López Bautista y Byron Humberto González Ramírez.

Este texto está protegido por la [Licencia Creative Commons Atribución 4.0 Internacional](#)



Este texto está protegido por una licencia
[Creative Commons 4.0.](#)

Es libre para compartir, copiar y redistribuir el material en cualquier medio o formato y adaptar el documento, remezclar, transformar y crear a partir del material para cualquier propósito, incluso comercialmente, siempre que cumpla la condición de atribución: debe reconocer el crédito de una obra de manera adecuada, proporcionar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que tiene el apoyo del licenciante o lo recibe por el uso que hace.