

Diseño de experimentos factoriales 2ⁿ para su análisis con Infostat e Infogen

Delfina de Jesús Pérez López
J. Ramón Pascual Franco Martínez
Francisco Gutiérrez Rodríguez
Jesús Hernández Ávila
Artemio Balbuena Melgarejo
Andrés González Huerta[§]

Centro de Investigación y Estudios Avanzados en Fitomejoramiento-Facultad de Ciencias Agrícolas-Universidad Autónoma del Estado de México. *Campus Universitario ‘El Cerrillo’*, Toluca, Estado de México, México. AP. 435. Tel. 7222965518, ext. 148. (djperezl@uaemex.mx; jrfrancom@uaemex.mx; fgutierrezr@uaemex.mx; jhernandez@uaemex.mx; abalbuenam@uaemex.mx).

[§]Autor para correspondencia: agonzalezh@uaemex.mx.

Resumen

El diseño y el análisis de datos de experimentos factoriales usando paquetes estadísticos ahorra tiempo, simplifica los cálculos y permite la discusión de la contribución individual o conjunta de varios factores. En este estudio se indican los procedimientos para analizar datos con los paquetes estadísticos InfoStat e InfoGen en experimentos factoriales 2ⁿ, con n= 2, 3, 4, 5, 6, 7 y 8 factores de la productividad. Además, se muestra el diseño de su estructura de tratamientos, las interacciones posibles, los modelos estadísticos para un diseño de bloques completos al azar y algunas fórmulas para calcular sumas de cuadrados para n= 8. Se establecen las relaciones entre los modelos lineales y los procedimientos que se deben aplicar en ambos paquetes estadísticos para obtener un análisis de varianza y la comparación de medias de tratamientos si se utiliza el modelo estadístico general, también se indican las instrucciones para estimar los efectos para cada uno de los ocho factores principales y para sus interacciones posibles con la prueba de Tukey ($p= 0.01$). El diseño de este tipo de experimentos se realizó para ser analizado con versiones académicas de prueba, las cuales son gratuitas por seis meses en ambos paquetes estadísticos. El usuario podría adquirir una licencia por un año para instalarla en dos computadoras personales pagando, por cada software, solamente USD \$50.00. Ambos paquetes estadísticos son muy amigables y pueden utilizarse para analizar cualquier diseño experimental, si se indica el modelo estadístico apropiado.

Palabras clave: bloques completos al azar, construcción de tratamientos factoriales, diseño experimental exploratorio, modelos estadísticos.

Recibido: junio de 2021

Aceptado: agosto de 2021

Introducción

La naturaleza secuencial e iterativa de la metodología de superficies de respuesta requiere que en cada fase se identifiquen los factores significativos, se realice una adecuación de modelo y se determine la dirección en la cual se encuentran las condiciones óptimas de experimentación. Para identificar los factores que tienen mayor influencia en la respuesta y para examinar la bondad del ajuste del modelo estadístico se debe recurrir al análisis de varianza (Jiménez, 2015).

En este contexto, el diseño y el análisis de experimentos factoriales en bloques completos al azar ha sido utilizado como una herramienta muy valiosa en las ciencias agropecuarias, y en otras ramas de la ciencia y de la tecnología (Sahagún, 1998; Medina y López, 2011; Jiménez, 2015; Álvarez *et al.*, 2018), además de determinar la contribución de cada uno de los factores estudiados permite simultáneamente saber cuáles son los efectos de sus interacciones, en cada ensayo y a través de años, localidades o sus combinaciones en tiempo y espacio (Montgomery, 2010; Gomez y Gomez, 1984; Di Rienzo *et al.*, 2008; Balzarini y Di Rienzo, 2016).

Los experimentos 2^n son familias de ensayos con n factores y dos niveles dentro de cada uno de ellos que generalmente representan intervalos de variabilidad muy amplios en estudios exploratorios, cuando no existe o cuando hay poca información publicada y consecuentemente, es deseable rediseñar su investigación con un subconjunto más importante (Gomez y Gomez, 1984; Montgomery, 2010; Medina y López, 2011; Jiménez, 2015).

La evaluación de dos cruzas promisorias formadas con las líneas más sobresalientes, la introducción a una nueva región de cultivos comerciales, de especies vegetales en proceso de domesticación o de animales de varias razas, así como la generación, validación, aplicación y transferencia de tecnología podría representar un caso deseable para el diseño y el análisis de este tipo de investigaciones (Torres *et al.*, 2011; Reynoso *et al.*, 2014; Torres *et al.*, 2017; González *et al.*, 2019). La determinación de dosis óptimas cuando se aplican fertilizantes orgánicos y/o inorgánicos, en diferentes densidades de población, en varias fechas de siembra, o diversas localidades y años, podría tener como un prerrequisito, el análisis de datos de ensayos exploratorios (Rodríguez *et al.*, 2015; Quiroz *et al.*, 2017; Padilla *et al.*, 2019; González *et al.*, 2019).

Los experimentos factoriales 2^n son difíciles de analizar con una calculadora de escritorio cuando n aumenta, debido a un incremento en el número de interacciones que deben estimarse, cuando $n= 6, 7$ y 8 hay $57, 120$ y 247 interacciones, respectivamente. Con el uso de paquetes estadísticos su diseño y su análisis ahorra tiempo y simplifica los cálculos, especialmente en series de experimentos o en arreglos de unidades experimentales que se diseñan para facilitar su manejo en campo y para obtener diferentes niveles de precisión en la estimación de los factores que los conforman (Gomez y Gomez, 1984; Martínez, 1988; Padilla *et al.*, 2019; González *et al.*, 2019).

Las primeras versiones de InfoStat e InfoGen aparecieron antes de 2008 y 2016, respectivamente, ambas fueron creadas por programadores y analistas de la Universidad de Córdoba, en Argentina y sus versiones académicas pueden descargarse gratuitamente desde sus sitios web.

Ambos softwares son muy amigables y pueden utilizarse para analizar cualquier tipo de experimento si se introduce el modelo estadístico correcto, debido a que éstos ahora pueden potenciar su uso; a través, de R-software (Sahagún, 1998; Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini y Di Rienzo, 2016). Como referencia en el uso del SAS (Statistical Analysis System, 1989); InfoStat e InfoGen; Padilla *et al.* (2019); González *et al.* (2019), analizaron datos de la producción en vaina verde de tres cultivares de haba (*Vicia faba* L.), sembrados en dos localidades mexiquenses, con tres densidades de población y tres fórmulas de fertilización: en la primera publicación se presentaron los códigos para analizar y validar una serie de experimentos en bloques completos al azar en arreglo de parcelas subdivididas y en la segunda publicación se describió el programa para estimar contrastes mutuamente ortogonales en ese tipo de ensayos.

En el contexto anterior, el objetivo principal de este estudio fue diseñar los procedimientos para analizar experimentos factoriales 2^n en bloques completos al azar con r repeticiones usando InfoStat e InfoGen, para generar un análisis de varianza y una comparación de medias de tratamientos con la prueba de Tukey.

Materiales y métodos

Notación utilizada

Los ocho factores serán representados como A, B, H; ambos niveles dentro de cada uno de éstos serán designados por la misma letra, en minúscula y con un subíndice, como a_1 y a_2 para el factor A. Las interacciones serán designadas como $A*B$, $A*C$, $A*B*C*D*E*F*G*H$. En los modelos estadísticos cada factor estará asociado a un subíndice: A, B, C, D, E, F, G, H tendrán correspondencia con i, j, k, l, m, n, o, p, respectivamente. Las repeticiones serán identificadas con R y el último subíndice que se indica en el residual de cada modelo utilizado (Sahagún, 1998; Balzarini *et al.*, 2008; Balzarini y Di Rienzo, 2016).

Número de tratamientos y su estructura factorial

Los tratamientos serán representados por la combinación de los niveles de dos o más factores. El diseño experimental es una estrategia de combinación de la estructura de tratamientos con la estructura de unidades experimentales de manera tal que las alteraciones en las respuestas, de al menos en algún subconjunto de unidades experimentales, pueda ser atribuida exclusivamente a la acción de los tratamientos, excepto por variaciones aleatorias. Así, es posible comparar medias de tratamientos o combinaciones lineales de medias de tratamientos con el menor ruido posible (Di Rienzo *et al.*, 2008; Balzarini *et al.*, 2016; Balzarini y Di Rienzo, 2016). En el Cuadro 1 se muestra la terminología, cuantos y cuál es la estructura de tratamientos para $n=2, 3, 4, 5, 6, 7, 8$.

Cuadro 1. Cantidad de factores, códigos para representarlos, número y estructura de tratamientos.

n	Simbología	Número	Tratamientos que serán analizados
2	A, B	4	$a_1b_1, a_1b_2, a_2b_1, a_2b_2$
3	A, B, C	8	$a_1b_1c_1, a_1b_1c_2, a_1b_2c_1, a_1b_2c_2, a_2b_1c_1, a_2b_1c_2, a_2b_2c_1, a_2b_2c_2$

n	Simbología	Número	Tratamientos que serán analizados
4	A, B, C, D	16	a ₁ b ₁ c ₁ d ₁ , a ₁ b ₁ c ₁ d ₂ , a ₁ b ₁ c ₂ d ₁ , a ₁ b ₁ c ₂ d ₂ , a ₁ b ₂ c ₁ d ₁ , a ₁ b ₂ c ₁ d ₂ , a ₁ b ₂ c ₂ d ₁ , a ₁ b ₂ c ₂ d ₂ , a ₂ b ₁ c ₁ d ₁ , a ₂ b ₁ c ₁ d ₂ , a ₂ b ₁ c ₂ d ₁ , a ₂ b ₁ c ₂ d ₂ , a ₂ b ₂ c ₁ d ₁ , a ₂ b ₂ c ₁ d ₂ , a ₂ b ₂ c ₂ d ₁ , a ₂ b ₂ c ₂ d ₂
5	A, B, C, D, E	32	a ₁ b ₁ c ₁ d ₁ e ₁ , a ₁ b ₁ c ₁ d ₁ e ₂ , a ₁ b ₁ c ₁ d ₂ e ₁ , a ₁ b ₁ c ₁ d ₂ e ₂ , a ₁ b ₁ c ₂ d ₁ e ₁ , a ₁ b ₁ c ₂ d ₁ e ₂ , a ₁ b ₁ c ₂ d ₂ e ₁ , a ₁ b ₁ c ₂ d ₂ e ₂ , a ₁ b ₂ c ₁ d ₁ e ₁ , a ₁ b ₂ c ₁ d ₁ e ₂ , a ₁ b ₂ c ₁ d ₂ e ₁ , a ₁ b ₂ c ₁ d ₂ e ₂ , a ₁ b ₂ c ₂ d ₁ e ₁ , a ₁ b ₂ c ₂ d ₁ e ₂ , a ₁ b ₂ c ₂ d ₂ e ₁ , a ₁ b ₂ c ₂ d ₂ e ₂ , a ₂ b ₁ c ₁ d ₁ e ₁ , a ₂ b ₁ c ₁ d ₁ e ₂ , a ₂ b ₁ c ₁ d ₂ e ₁ , a ₂ b ₁ c ₁ d ₂ e ₂ , a ₂ b ₁ c ₂ d ₁ e ₁ , a ₂ b ₁ c ₂ d ₁ e ₂ , a ₂ b ₁ c ₂ d ₂ e ₁ , a ₂ b ₁ c ₂ d ₂ e ₂ , a ₂ b ₂ c ₁ d ₁ e ₁ , a ₂ b ₂ c ₁ d ₁ e ₂ , a ₂ b ₂ c ₁ d ₂ e ₁ , a ₂ b ₂ c ₁ d ₂ e ₂ , a ₂ b ₂ c ₂ d ₁ e ₁ , a ₂ b ₂ c ₂ d ₁ e ₂ , a ₂ b ₂ c ₂ d ₂ e ₁ , a ₂ b ₂ c ₂ d ₂ e ₂
6	A, B, C, D, E, F	64	a ₁ b ₁ c ₁ d ₁ e ₁ f ₁ , a ₁ b ₁ c ₁ d ₁ e ₁ f ₂ , a ₁ b ₁ c ₁ d ₁ e ₂ f ₁ , a ₁ b ₁ c ₁ d ₁ e ₂ f ₂ , a ₁ b ₁ c ₁ d ₂ e ₁ f ₁ , a ₁ b ₁ c ₁ d ₂ e ₁ f ₂ , a ₁ b ₁ c ₁ d ₂ e ₂ f ₁ , a ₁ b ₁ c ₁ d ₂ e ₂ f ₂ , a ₁ b ₁ c ₂ d ₁ e ₁ f ₁ , a ₁ b ₁ c ₂ d ₁ e ₁ f ₂ , a ₁ b ₁ c ₂ d ₂ e ₁ f ₁ , a ₁ b ₁ c ₂ d ₂ e ₁ f ₂ , a ₁ b ₁ c ₂ d ₂ e ₂ f ₁ , a ₁ b ₁ c ₂ d ₂ e ₂ f ₂ , a ₁ b ₂ c ₁ d ₁ e ₁ f ₁ , a ₁ b ₂ c ₁ d ₁ e ₁ f ₂ , a ₁ b ₂ c ₁ d ₂ e ₁ f ₁ , a ₁ b ₂ c ₁ d ₂ e ₁ f ₂ , a ₁ b ₂ c ₁ d ₂ e ₂ f ₁ , a ₁ b ₂ c ₁ d ₂ e ₂ f ₂ , a ₁ b ₂ c ₂ d ₁ e ₁ f ₁ , a ₁ b ₂ c ₂ d ₁ e ₁ f ₂ , a ₁ b ₂ c ₂ d ₁ e ₂ f ₁ , a ₁ b ₂ c ₂ d ₁ e ₂ f ₂ , a ₁ b ₂ c ₂ d ₂ e ₁ f ₁ , a ₁ b ₂ c ₂ d ₂ e ₁ f ₂ , a ₁ b ₂ c ₂ d ₂ e ₂ f ₁ , a ₁ b ₂ c ₂ d ₂ e ₂ f ₂ , a ₂ b ₁ c ₁ d ₁ e ₁ f ₁ , a ₂ b ₁ c ₁ d ₁ e ₂ f ₁ , a ₂ b ₁ c ₁ d ₂ e ₁ f ₁ , a ₂ b ₁ c ₁ d ₂ e ₂ f ₁ , a ₂ b ₁ c ₂ d ₁ e ₁ f ₁ , a ₂ b ₁ c ₂ d ₁ e ₂ f ₁ , a ₂ b ₁ c ₂ d ₂ e ₁ f ₁ , a ₂ b ₁ c ₂ d ₂ e ₂ f ₁ , a ₂ b ₂ c ₁ d ₁ e ₁ f ₁ , a ₂ b ₂ c ₁ d ₁ e ₂ f ₁ , a ₂ b ₂ c ₁ d ₂ e ₁ f ₁ , a ₂ b ₂ c ₁ d ₂ e ₂ f ₁ , a ₂ b ₂ c ₂ d ₁ e ₁ f ₁ , a ₂ b ₂ c ₂ d ₁ e ₂ f ₁ , a ₂ b ₂ c ₂ d ₂ e ₁ f ₁ , a ₂ b ₂ c ₂ d ₂ e ₂ f ₁ , a ₂ b ₂ c ₂ d ₂ e ₂ f ₂ , a ₂ b ₂ c ₂ d ₂ e ₂ f ₂
7	A, B, C, D, E, F, G	128	a ₁ b ₁ c ₁ d ₁ e ₁ f ₁ g ₁ , a ₁ b ₁ c ₁ d ₁ e ₁ f ₁ g ₂ , a ₁ b ₁ c ₁ d ₁ e ₁ f ₂ g ₁ , a ₁ b ₁ c ₁ d ₁ e ₁ f ₂ g ₂ , a ₁ b ₁ c ₁ d ₁ e ₂ f ₁ g ₁ , a ₁ b ₁ c ₁ d ₁ e ₂ f ₁ g ₂ , a ₁ b ₁ c ₁ d ₁ e ₂ f ₂ g ₁ , a ₁ b ₁ c ₁ d ₁ e ₂ f ₂ g ₂ , a ₁ b ₁ c ₁ d ₂ e ₁ f ₁ g ₁ , a ₁ b ₁ c ₁ d ₂ e ₁ f ₁ g ₂ , a ₁ b ₁ c ₁ d ₂ e ₁ f ₂ g ₁ , a ₁ b ₁ c ₁ d ₂ e ₁ f ₂ g ₂ , a ₁ b ₁ c ₁ d ₂ e ₂ f ₁ g ₁ , a ₁ b ₁ c ₁ d ₂ e ₂ f ₁ g ₂ , a ₁ b ₁ c ₁ d ₂ e ₂ f ₂ g ₁ , a ₁ b ₁ c ₁ d ₂ e ₂ f ₂ g ₂ , a ₁ b ₁ c ₂ d ₁ e ₁ f ₁ g ₁ , a ₁ b ₁ c ₂ d ₁ e ₁ f ₁ g ₂ , a ₁ b ₁ c ₂ d ₁ e ₁ f ₂ g ₁ , a ₁ b ₁ c ₂ d ₁ e ₁ f ₂ g ₂ , a ₁ b ₁ c ₂ d ₁ e ₂ f ₁ g ₁ , a ₁ b ₁ c ₂ d ₁ e ₂ f ₁ g ₂ , a ₁ b ₁ c ₂ d ₁ e ₂ f ₂ g ₁ , a ₁ b ₁ c ₂ d ₁ e ₂ f ₂ g ₂ , a ₁ b ₁ c ₂ d ₂ e ₁ f ₁ g ₁ , a ₁ b ₁ c ₂ d ₂ e ₁ f ₁ g ₂ , a ₁ b ₁ c ₂ d ₂ e ₁ f ₂ g ₁ , a ₁ b ₁ c ₂ d ₂ e ₁ f ₂ g ₂ , a ₁ b ₁ c ₂ d ₂ e ₂ f ₁ g ₁ , a ₁ b ₁ c ₂ d ₂ e ₂ f ₁ g ₂ , a ₁ b ₁ c ₂ d ₂ e ₂ f ₂ g ₁ , a ₁ b ₁ c ₂ d ₂ e ₂ f ₂ g ₂

Modelos estadísticos

Para construir los modelos estadísticos utilizados en el análisis de datos, sólo deberán elegirse las letras mayúsculas usadas para representar a los factores principales y las correspondientes a sus interacciones posibles, a un valor específico de n , después se eligen los subíndices que los identifican, y se adiciona μ , R y ε . Su construcción se basó en los procedimientos descritos por Sahagún (1998). Por ejemplo, en un experimento factorial 2^3 , si Y es la variable cuantitativa de interés: $Y_{ijkl} = \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + R_l + \varepsilon_{ijkl}$.

Donde: μ es la media aritmética de los abcr datos; A_i , B_j y C_k son los efectos del i-ésimo, j-ésimo o k-ésimo nivel de A, B o C, respectivamente; $(AB)_{ij}$, $(AC)_{ik}$, $(BC)_{jk}$ y $(ABC)_{ijk}$ son las interacciones entre dos o tres factores; R_l es el efecto de la l-ésima repetición y ε_{ijkl} es el residual del modelo. En el Cuadro 2 se muestran los modelos estadísticos que serán utilizados para analizar los datos cuando $n = 2, 3, 8$.

Cuadro 2. Modelos estadísticos usados en el análisis de datos con InfoStat e InfoGen.

Valor de n	Modelo estadístico
2	$Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + R_k + \varepsilon_{ijk}$
3	$Y_{ijkl} = \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik} + (BC)_{jk} + (ABC)_{ijk} + R_l + \varepsilon_{ijkl}$
4	$Y_{ijklm} = \mu + A_i + B_j + C_k + D_l + (AB)_{ij} + (AC)_{ik} + (AD)_{il} + (BC)_{jk} + (BD)_{jl} + (CD)_{kl} + (ABC)_{ijk} + (ABD)_{ijl} + (ACD)_{ikl} + (BCD)_{jkl} + (ABCD)_{ijkl} + R_m + \varepsilon_{ijklm}$
5	$Y_{ijklmn} = \mu + A_i + B_j + C_k + D_l + E_m + \{ \text{interacciones entre dos, tres, cuatro y cinco factores} \}^{\&} + R_n + \varepsilon_{ijklmn}$
6	$Y_{ijklmno} = \mu + A_i + B_j + C_k + D_l + E_m + F_n + \{ \text{interacciones entre dos, tres, cuatro, cinco y seis factores} \}^{\&} + R_o + \varepsilon_{ijklmno}$
7	$Y_{ijklmnop} = \mu + A_i + B_j + C_k + D_l + E_m + F_n + G_o + \{ \text{interacciones entre dos, tres, cuatro, cinco, seis y siete factores} \}^{\&} + R_p + \varepsilon_{ijklmnop}$
8	$Y_{ijklmnopq} = \mu + A_i + B_j + C_k + D_l + E_m + F_n + G_o + H_p + \{ \text{interacciones entre dos, tres, cuatro, cinco, seis, siete y ocho factores} \}^{\&} + R_q + \varepsilon_{ijklmnopq}$
2 a 8	$Y_{ij} = \mu + \tau_i + R_j + \varepsilon_{ij}$ Donde: $\tau_i = 2^2, 2^3, 2^4, 2^5, 2^6, 2^7, \text{ ó } 2^8$ tratamientos

[&]= las interacciones posibles para n=5, 6, 7 y 8 se muestran posteriormente.

Para calcular cuantas interacciones serán indicadas en los modelos estadísticos simplemente se suman las combinaciones que son posibles para los n factores tomados k veces, tal que $2 \leq k \leq 8$; k representa, en ese orden, interacciones entre dos, tres, ocho factores. Si n= 8: $C_2^8 + C_3^8 + C_4^8 + C_5^8 + C_6^8 + C_7^8 + C_8^8 = 28 + 56 + 70 + 56 + 28 + 8 + 1 = 247$ interacciones. Para verificar: $2^n - C_0^n - C_1^n = 2^8 - C_0^8 - C_1^8 = 256 - 1 - 8 = 247$ interacciones.

Interacciones para n= 8

Entre dos factores: A*B[£], A*C[£], A*D[£], A*E[£], A*F, A*G, A*H, B*C[£], B*D[£], B*E[£], B*F, B*G, B*H, C*D[£], C*E[£], C*F, C*G, C*H, D*E[£], D*F, D*G, D*H, E*F, E*G, E*H, F*G, F*H, G*H.

Entre tres factores: A*B*C[£], A*B*D[£], A*B*E[£], A*B*F, A*B*G, A*B*H, A*C*D[£], A*C*E[£], A*C*F, A*C*G, A*C*H, A*D*E[£], A*D*F, A*D*G, A*D*H, A*E*F, A*E*G, A*E*H, A*F*G, A*F*H, A*G*H, B*C*D[£], B*C*E[£], B*C*F, B*C*G, B*C*H, B*D*E[£], B*D*F, B*D*G, B*D*H, B*E*F, B*E*G, B*E*H, B*F*G, B*F*H, B*G*H, C*D*E[£], C*D*F, C*D*G, C*D*H, C*E*F, C*E*G, C*E*H, C*F*G, C*F*H, C*G*H, D*E*F, D*E*G, D*E*H, D*F*G, D*F*H, D*G*H, E*F*G, E*F*H, E*G*H, F*G*H.

Entre cuatro factores: A*B*C*D[£], A*B*C*E[£], A*B*C*F, A*B*C*G, A*B*C*H, A*B*D*E[£], A*B*D*F, A*B*D*G, A*B*D*H, A*B*E*F, A*B*E*G, A*B*E*H, A*B*F*G, A*B*F*H, A*B*G*H, A*C*D*E[£], A*C*D*F, A*C*D*G, A*C*D*H, A*C*E*F, A*C*E*G, A*C*E*H, A*C*F*G, A*C*F*H, A*C*G*H, A*D*E*F, A*D*E*G, A*D*E*H, A*D*F*G, A*D*F*H, A*D*G*H, A*E*F*G, A*E*F*H, A*E*G*H, A*F*G*H, B*C*D*E[£], B*C*D*F, B*C*D*G,

B*C*D*H, B*C*E*F, B*C*E*G, B*C*E*H, B*C*F*G, B*C*F*H, B*C*G*H, B*D*E*F, B*D*E*G, B*D*E*H, B*D*F*G, B*D*F*H, B*D*G*H, B*E*F*G, B*E*F*H, B*E*G*H, B*F*G*H, C*D*E*F, C*D*E*G, C*D*E*H, C*D*F*G, C*D*F*H, C*D*G*H, C*E*F*G, C*E*F*H, C*E*G*H, C*F*G*H, D*E*F*G, D*E*F*H, D*E*G*H, D*F*G*H, E*F*G*H.

Entre cinco factores: A*B*C*D*E^f, A*B*C*D*F, A*B*C*D*G, A*B*C*D*H, A*B*C*E*F, A*B*C*E*G, A*B*C*E*H, A*B*C*F*G, A*B*C*F*H, A*B*C*G*H, A*B*D*E*F, A*B*D*E*G, A*B*D*E*H, A*B*D*F*G, A*B*D*F*H, A*B*D*G*H, A*B*E*F*G, A*B*E*F*H, A*B*E*G*H, A*B*F*G*H, A*C*D*E*F, A*C*D*E*G, A*C*D*E*H, A*C*D*F*G, A*C*D*F*H, A*C*D*G*H, A*C*E*F*G, A*C*E*F*H, A*C*E*G*H, A*C*F*G*H, A*D*E*F*G, A*D*E*F*H, A*D*E*G*H, A*D*F*G*H, A*E*F*G*H, B*C*D*E*F, B*C*D*E*G, B*C*D*E*H, B*C*D*F*G, B*C*D*F*H, B*C*D*G*H, B*C*E*F*G, B*C*E*F*H, B*C*E*G*H, B*C*F*G*H, B*D*E*G*H, B*D*F*G*H, B*E*F*G*H, C*D*E*F*G, C*D*E*F*H, C*D*E*G*H, C*D*F*G*H, C*E*F*G*H, D*E*F*G*H.

Entre seis factores: A*B*C*D*E*F, A*B*C*D*E*G, A*B*C*D*E*H, A*B*C*D*F*G, A*B*C*D*F*H, A*B*C*D*G*H, A*B*C*E*F*G, A*B*C*E*F*H, A*B*C*E*G*H, A*B*C*F*G*H, A*B*D*E*F*G, A*B*D*E*F*H, A*B*D*E*G*H, A*B*D*F*G*H, A*B*E*F*G*H, A*C*D*E*F*G, A*C*D*E*F*H, A*C*D*E*G*H, A*C*D*F*G*H, A*C*E*F*G*H, A*D*E*F*G*H, B*C*D*E*F*G, B*C*D*E*F*H, B*C*D*E*G*H, B*C*D*F*G*H, B*C*D*F*H, B*C*D*G*H, B*C*D*F*G*H, B*C*E*F*G*H, B*D*E*F*G*H, B*E*F*G*H, C*D*E*F*G, C*D*E*F*H, C*D*E*G*H.

Entre siete factores: A*B*C*D*E*F*G, A*B*C*D*E*F*H, A*B*C*D*E*G*H, A*B*C*D*F*G*H, A*B*C*E*F*G*H, A*B*D*E*F*G*H, A*B*D*E*F*H, A*C*D*E*F*G*H, B*C*D*E*F*G*H.

Entre ocho factores: A*B*C*D*E*F*G*H

Las interacciones anteriores también pueden utilizarse para construir cualquier modelo estadístico, como los correspondientes a las series de experimentos en bloques completos al azar en arreglo de parcelas subdivididas (Padilla *et al.*, 2019; González *et al.*, 2019). Por ejemplo, las 26 interacciones que se originan entre los factores A, B, C, D y E, y que deben alimentarse en especificaciones al modelo de InfoStat e InfoGen, fueron identificadas previamente con el superíndice *f*.

Algunas fórmulas para verificar sumas de cuadrados con n= 8

Las fórmulas presentadas a continuación fueron obtenidas con los procedimientos descritos por Sahagún (1998).

$$\begin{aligned} \text{SC Factor A} &= \frac{\sum_{i=1}^a Y_i^2}{bcdefghr} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefgahr} \\ \text{SC Factor B} &= \frac{\sum_{j=1}^b Y_j^2}{acdefghr} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefgahr} \\ \text{SC Factor C} &= \frac{\sum_{k=1}^c Y_k^2}{abdefghr} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefgahr} \end{aligned}$$

$$\text{SC Factor D} = \frac{\sum_{l=1}^d Y_{ijklmnopq}^2}{abcdefg} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefg}$$

$$\text{SC Factor E} = \frac{\sum_{m=1}^e Y_{ijklmnopq}^2}{abcdefg} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefg}$$

$$\text{SC Factor F} = \frac{\sum_{n=1}^f Y_{ijklmnopq}^2}{abcdefg} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefg}$$

$$\text{SC Factor G} = \frac{\sum_{o=1}^g Y_{ijklmnopq}^2}{abcdefg} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefg}$$

$$\text{SC Factor H} = \frac{\sum_{p=1}^h Y_{ijklmnopq}^2}{abcdefg} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefg}$$

$$\text{SC Repeticiones} = \frac{\sum_{q=1}^r Y_{ijklmnopq}^2}{abcdefg} - \frac{(\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h \sum_{q=1}^r Y_{ijklmnopq})^2}{abcdefg}$$

A continuación, la segunda parte de cada fórmula, que corresponde al factor de corrección, será identificada como FC y por conveniencia, serán definidas las SC de tratamientos 1, 2, s para que por diferencia sea calculada la SC de la interacción de interés. El procedimiento es similar al reportado en Padilla *et al.* (2019), para una serie de experimentos en bloques completos al azar en arreglo de parcelas subdivididas.

SC para una interacción entre dos factores

$$\text{SC Tratamientos 1} = \text{SC A} + \text{SC B} + \text{SC A*B}$$

$$\text{SC A*B} = \text{SC Tratamientos 1} - \text{SC A} - \text{SC B}$$

$$\text{Donde: SC Tratamientos 1} = \frac{\sum_{i=1}^a \sum_{j=1}^b Y_{ij}^2}{cd} - \text{FC}$$

SC para una interacción entre tres factores

$$\text{SC A*B*C} = \text{SC Tratamientos 2} - \text{SC A} - \text{SC B} - \text{SC C} - \text{SC A*B} - \text{SC A*C} - \text{SC B*C}$$

$$\text{Donde: SC Tratamientos 2} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c Y_{ijk}^2}{def} - \text{FC}$$

SC para una interacción entre cuatro factores

$$\text{SC A*B*C*D} = \text{SC Tratamientos 3} - \text{SC A} - \text{SC B} - \text{SC C} - \text{SC D} - \text{SC A*B} - \text{SC A*C} - \text{SC A*D} - \text{SC B*C} - \text{SC B*D} - \text{SC C*D} - \text{SC A*B*C} - \text{SC A*B*D} - \text{SC A*C*D}. \text{ Dónde: SC Tratamientos}$$

$$3 = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d Y_{ijkl}^2}{efgh} - \text{FC}.$$

SC para una interacción entre cinco factores

$$\text{SC A*B*C*D*E} = \text{SC Tratamientos 4} - \text{SC A} - \text{SC B} - \text{SC C} - \text{SC D} - \text{SC E} - (\text{SC interacciones}$$

entre dos, tres y cuatro factores). Donde: SC Tratamientos 4 = $\frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e Y_{ijklm}^2}{fg} - \text{FC}$. En lo sucesivo, las SC para A, B, C, D, E, F, G, H serán identificadas como SC factores principales.

SC para una interacción de seis factores

$SC_{A*B*C*D*E*F} = SC_{\text{Tratamientos } 5} - SC_{\text{factores principales}} - SC_{\text{interacciones entre dos, tres, cuatro y cinco factores}}$. Donde: $SC_{\text{Tratamientos } 5} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f Y_{ijklmn}^2}{ghr} - FC$.

SC para una interacción entre siete factores

$SC_{A*B*C*D*E*F*G} = SC_{\text{Tratamientos } 6} - SC_{\text{factores principales}} - SC_{\text{interacciones entre dos, tres, cuatro, cinco y seis factores}}$. Donde: $SC_{\text{Tratamientos } 6} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g Y_{ijklmon}^2}{hr} - FC$.

SC para la interacción entre ocho factores

$SC_{A*B*C*D*E*F*G*H} = SC_{\text{Tratamientos } 7} - SC_{\text{factores principales}} - SC_{\text{interacciones entre dos, tres, cuatro, cinco, seis y siete factores}}$. Donde: $SC_{\text{Tratamientos } 7} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c \sum_{l=1}^d \sum_{m=1}^e \sum_{n=1}^f \sum_{o=1}^g \sum_{p=1}^h Y_{ijklmop}^2}{r} - FC$.

Elaboración de bases de datos

Éstas deberán tener la estructura del Cuadro 3, en caso se observa el número parcial o total de datos capturados, Rep identifica repeticiones, Trat define al tratamiento, Ren especifica el rendimiento y en A, B, H se registran las combinaciones entre ambos niveles de cada uno de los ocho factores.

Cuadro 3. Estructura de la base de datos de un experimento factorial 2⁸.

Caso	Rep	Trat	A	B	C	D	E	F	G	H	Ren
1	1	01	1	1	1	1	1	1	1	1	4
2	1	02	1	1	1	1	1	1	1	2	6.3
3	1	03	1	1	1	1	1	1	2	1	5.3
4	1	04	1	1	1	1	1	1	2	2	6.9
5	1	05	1	1	1	1	1	2	1	1	6.5
6	1	06	1	1	1	1	1	2	1	2	7.9
7	1	07	1	1	1	1	1	2	2	1	7.2
8	1	08	1	1	1	1	1	2	2	2	8.3
.
.
.
m-1	r	t-1	a-1	b-1	c-1	d-1	e-1	f-1	g-1	h-1	8.6
m	r	t	a	b	c	d	e	f	g	h	9.1

$m = 2^n r$ observaciones, r = número de repeticiones; Rep= repetición; Trat= tratamiento; Ren= rendimiento ($t \text{ ha}^{-1}$).

El resguardo de esta estructura de tratamientos sigue las mismas recomendaciones que para cualquier archivo que contenga bases de datos, el archivo podría llamarse EXP2N.IDB2.

Procedimiento para analizar los datos

Después de cargar InfoStat o InfoGen, en la pantalla se muestran dos cuadros de diálogo, en ambos elegir ok, aparecerá su menú principal. En archivo elegir la opción abrir para cargar la base de datos. InfoStat o InfoGen mostrará una tabla similar a la del Cuadro 3.

En estadísticas elegir análisis de la varianza y aparecerá otro cuadro de diálogo donde se solicita la definición de las variables dependientes y de clasificación. Si el objetivo es analizar el experimento factorial 2^8 , la variable dependiente es rinde, las variables de clasificación son Rep, A, B, C, D, E, F, G, H. Dar clic en aceptar.

En la pantalla se mostrará otro cuadro de diálogo que dice análisis de varianza. Aquí, el usuario verificará que en especificaciones de los términos del modelo sean correctas las variables de clasificación (Rep, A, B, H), las interacciones A*B, A*C, A*B*C*D*E*F*G*H deben introducirse debajo del último factor principal. En un experimento factorial 2^3 , en el cuadro de diálogo debe mostrarse verticalmente Rep, A, B, C, A*B, A*C, B*C, A*B*C. No es necesario especificar ni la media aritmética ni el residual del modelo. Se obtendrá el análisis de varianza.

En el último cuadro de diálogo podría elegirse una comparación de medias de tratamientos con LSD Fisher, Bonferroni, Tukey, Duncan y Scheffé, entre otras. En comparaciones elegir: Tukey/mostrar medias según: Rep, A, B, C, H, A*B, A*C, A*B*C*D*E*F*G*H/ presentación en lista descendente/nivel de significancia 0.01/aceptar. Este procedimiento genera el análisis de varianza y la comparación de medias con la prueba de Tukey ($p= 0.01$) para los ocho factores y para sus 247 interacciones.

Conclusiones

Si el usuario quiere realizar un análisis de varianza usando el modelo estadístico general y una comparación de medias, en el cuadro de diálogo correspondiente, en variables de clasificación debe elegir rep trat y rinde. InfoStat e InfoGen sólo mostrará la salida con los 256 tratamientos para un diseño de bloques completos al azar, si en variables de clasificación no se elige rep, entonces se obtendrá el análisis de varianza de un diseño completamente al azar.

Literatura citada

- Álvarez, G. L.; Guerra, S. J. y Rojas, R. N. 2018. Diseño factorial 2^k aplicado a la caracterización reológica de suspensiones de caolín. Prospectiva. 16(1):18-25.
- Balzarini, M. G.; González, L.; Tablada, M.; Casanoves, F.; Di Rienzo, J. A. y Robledo, C. W. 2008. InfoStat, versión 2008. Manual del usuario. Editorial Brujas, Córdoba, Argentina. 336 p.
- Balzarini, M.; Bruno, C. y Arroyo, A. 2016. Análisis de ensayos agrícolas multi-ambientales. ejemplos con Info-Gen. Manual del Usuario. Unidad de Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba, Argentina. 141 p.

- Balzarini, M. G. y Di Rienzo, J. A. 2016. InfoGen, versión 2016. FCA. Universidad Nacional de Córdoba, Argentina. www.info-gen.com.ar.
- Di Rienzo, J. A.; Casanoves, F.; Balzarini, M. G.; González, L.; Tablada, M. y Robledo, C. W. 2008. InfoStat, versión 2008. Grupo InfoStat, FCA. Universidad Nacional de Córdoba, Argentina. 336 p.
- Gomez, K. A. and Gomez, A., A. 1984. Statistical procedures for agricultural research. 2nd (Ed). John Wiley and Sons, Inc. Printed in Singapore. 680 p.
- González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R., F.; Franco, M. J. R.P. y Padilla, L. A. 2019. InfoStat, InfoGen y SAS para contrastes mutuamente ortogonales en experimentos en bloques completos al azar en parcelas subdivididas. *Rev. Mex. Cienc. Agríc.* 10(6):1417-1431.
- Jiménez, C. M. G. 2015. Superficies de respuesta mediante un diseño central compuesto. Instituto de Estadística Teórica y Aplicada. *Rev. Varianza*. 11(1):31-36.
- Martínez, G. A. 1988. Diseño de experimentos. Teoría y métodos. Editorial Trillas, 1^{ra} (Ed). México, DF. 756 p.
- Medina, V. P. D. y López, R. A. M. 2011. Análisis crítico del diseño factorial 2^k sobre casos aplicados. *Scientia Et Technica* XVII. 47(1):101-106.
- Montgomery, D. C. 2010. Diseño y análisis de experimentos. Editorial Limusa SA de CV. Grupo Noriega Editores. 2nd (Ed). México, DF. 686 p.
- Padilla, L. A.; González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R. F.; Ramírez, D. J. F.; Franco, M. J. R. P. y Serrato, C. R. 2019. Programas para SAS e InfoStat para analizar una serie de experimentos en parcelas subdivididas. Temas Selectos en la Innovación de las Ciencias Agropecuarias. 1. Agricultura. 2. Desarrollo Rural. Alfaomega Grupo Editor S.A. de C.V. 1^{ra} (Ed). México, DF. 67-88 pp. (ISBN:978-607-538-411-5).
- Quiroz, M. J.; Pérez, L. D. J.; González, H. A.; Rubí, A. M.; Gutiérrez, R. F.; Franco, M. J. R. P. y Ramírez, D. J. F. 2017. Respuesta de 10 cultivares de maíz a la densidad de población en tres localidades del centro mexiquense. *Rev. Mex. Cienc. Agríc.* 8(7):1521-1535.
- Reynoso, Q. C. A.; González, H. A.; Pérez, L. D. J.; Franco, M. O.; Torres, F. J. L.; Velázquez, C. G. A.; Breton, L. C.; Balbuena, M. A. y Mercado, V. O. 2014. Análisis de 17 híbridos de maíz sembrados en 17 ambientes de los Valles Altos del centro de México. *Rev. Mex. Cienc. Agríc.* 5(5):871-882.
- Rodríguez, F. I.; González, H. A.; Pérez, L. D. J. y Rubí, A. M. 2015. Efecto de cinco densidades de población en ocho cultivares de maíz sembrados en tres localidades del Valle de Toluca, México. *Rev. Mex. Cienc. Agríc.* 6(8):1943-1955.
- Sahagún, C. J. 1998. Construcción y análisis de los modelos fijos, aleatorios y mixtos. Universidad Autónoma Chapingo (UACH)-Departamento de Fitotecnia. Boletín técnico núm. 2. 65 p.
- SAS. 1989. Institute Inc. SAS/STAT User's Guide, version 6. Fourth edition, Volume 1, Cary, NC. USA. 943 p.
- Torres, F. J. L.; Morales, R. E. J.; González, H. A.; Laguna, C. A. y Córdova, O. H. 2011. Respuesta de híbridos trilineales y probadores de maíz en Valles Altos del centro de México. *Rev. Mex. Cienc. Agríc.* 2(6):829-844.
- Torres, F. J. L.; Mendoza, G. B.; Prassana, B. M.; Alvarado, G.; San Vicente, E. M. and Crossa, J. 2017. Grain yield and stability of white early maize hybrids in the highland valleys of Mexico. *Crop Sci.* 57(6):3002-3015.