

Tratamientos anidados dentro de un arreglo en grupos de bloques completos balanceados

Andrés González Huerta^{1,§}
Delfina de Jesús Pérez López¹
Jesús Hernández Ávila¹
J. Ramón Pascual Franco Martínez¹
Martín Rubí Arriaga¹
Artemio Balbuena Melgarejo¹

1 Facultad de Ciencias Agrícolas-Centro de Investigación y Estudios Avanzados en Fitomejoramiento-Universidad Autónoma del Estado de México. Campus Universitario 'El Cerrillo'. El Cerrillo Piedras Blancas, Toluca, Estado de México, México. AP. 435. Tel. 722 2965518, ext. 148. (djperetzl@uaemex.mx; jhernandez@uaemex.mx; jrfrancom@uaemex.mx; mrubia@uaemex.mx; abalbuenam@uaemex.mx).

Autor para correspondencia: agonzalezh@uaemex.mx

Resumen

Cuando se diseña y analiza un experimento o una serie de ensayos en tiempo o espacio podría ser de gran relevancia realizar una subdivisión del número de tratamientos por medio de la formación de grupos en los que se considere alguna diferencia importante entre éstos y alguna similitud dentro de ellos. En este estudio se analizó el caso presentado por Gomez y Gomez (1984) con relación al rendimiento de grano registrado en 45 variedades de arroz, clasificadas en tres grupos, se presentó su modelo estadístico para un diseño experimental de bloques completos al azar, se incluyeron fórmulas complementarias para calcular las sumas de cuadrados con las metodologías de mínimos cuadrados y formas cuadráticas o matriciales y se propone el procedimiento para generar una salida si fuera aplicado InfoStat. Adicionalmente, son mencionadas otras formas para calcular grados de libertad si el área experimental es dividida en unidad principal y subunidad, así como los correspondientes al residual del modelo o error b, las cuales simplifican los cálculos manuales. Se homologaron las fórmulas para ambas metodologías partiendo del uso formal de la simbología utilizada en las notaciones suma y punto, con base en éstas últimas se presentan sus formas cuadráticas. Se discute la diferencia entre un análisis de varianza convencional y el que es considerado en este trabajo, con base en las sumas de cuadrados finalmente, se indica cómo aplicar la prueba de Tukey para la comparación de medias de variedades dentro de cada grupo si es utilizado InfoStat. También se recomienda el uso de una calculadora de matrices para solucionar cálculos cuando se utilizan formas cuadráticas, la cual se encuentra disponible gratuitamente en su sitio WEB.

Palabras clave:

diseño de bloques completos al azar, formas cuadráticas, InfoStat, mínimos cuadrados.

Introducción

El diseño, el análisis y la interpretación de los datos que proporciona un experimento es una herramienta muy valiosa en la investigación agronómica cuando las inferencias estadísticas se fundamentan con información confiable (Sahagún, 1998; Sahagún, 2007; Montgomery, 2009). Las etapas generales consideradas durante éstos incluyen definición del problema, justificación del ensayo, planteamiento y análisis crítico de los objetivos e hipótesis científicas, estructura de tratamientos, número de repeticiones y tamaño de la unidad experimental, elección del diseño experimental, control local de la heterogeneidad asociada a las parcelas adyacentes o entre las que reciban un mismo tratamiento, naturaleza y tipo de variable registrada, material e instrumental utilizado, metodologías estadísticas aplicables, protocolo de investigación por aplicar, discusión de resultados, derivación de conclusiones e informe final (Gomez y Gomez, 1984; Martínez, 1988; Little y Hills, 2008).

También deberá considerarse el modelo estadístico apropiado, tipo de factor considerado (fijo, aleatorio o mixto), existencia o ausencia de ortogonalidad, relaciones de anidamiento o cruzamiento entre los factores estudiados y en sus interacciones, dependencia entre cuadrados medios y sus esperanzas matemáticas con las pruebas de hipótesis pertinentes, software utilizado, e inferencia estadística, entre otros (Sahagún, 1998; Piepho *et al.*, 2003; Restrepo, 2007a, 2007b). Además, deberá ponderarse el hecho de realizar los análisis sin y con submuestreo dentro de las unidades experimentales y justificar el caso, cuando eligiendo un tamaño de muestra pequeño, se decida o no, estimar efectos en lugar de varianzas (Zamudio y Alvarado, 1996; Cochran y Cox, 2004; González *et al.*, 2023).

En el presente estudio se analizará el caso presentado por Gomez y Gomez (1984), quienes agruparon 45 variedades de arroz (*Oriza sativa* L.) en tres grupos, cada uno con 15 materiales genéticos, en el diseño experimental de bloques completos al azar que ellos utilizaron, en arreglo de bloques completos balanceados, utilizaron tres repeticiones por tratamiento para evaluar el rendimiento de grano.

Ellos no presentaron su modelo estadístico pero la variabilidad total que fue estimada en esta característica cuantitativa fue fraccionada en diferencias entre grupos, entre repeticiones, error a, entre tratamientos dentro de cada grupo y error b o residual del modelo. Como no se dispone de permiso para usar los datos, sólo se proporcionará información adicional para el análisis de este tipo de arreglo de unidades experimentales, con énfasis en su modelo estadístico, fórmulas alternativas para calcular suma de cuadrados con dos metodologías y procedimiento que generará una salida usando InfoStat, para un análisis de varianza y una comparación de medias de variedades dentro de grupos aplicando la prueba de Tukey.

Materiales y métodos

Modelo estadístico

Con base en la guía proporcionada por Sahagún (1998); Piepho *et al.* (2003); Restrepo (2007), entre otros, el modelo que se construyó fue: $X_{ijk} = \mu + G_i + R_j + (GR)_{ij} + \tau_{k(i)} + \epsilon_{ijk}$. Donde: X es el rendimiento de grano en arroz o cualquier otra variable cuantitativa de interés; μ es la media aritmética de los rt datos; G_i es el efecto causado por el i -ésimo grupo; R_j es la contribución de la j -ésima repetición; $(GR)_{ij}$ es la interacción causada por el i -ésimo grupo con la j -ésima repetición o error a; $\tau_{k(i)}$ es el efecto causado por el k -ésimo tratamiento anidado dentro del i -ésimo grupo; y ϵ_{ijk} es el residual del modelo o error b.



Simbología utilizada para calcular suma de cuadrados

Las variables de clasificación en el modelo construido previamente son grupos, repeticiones y tratamientos, identificadas con i, j, k , sus niveles son $g, r, t/g$, respectivamente. En el presente estudio siempre $g=$ y ambos serán equivalentes a s , éste último usado por Gomez y Gomez (1984). Los tratamientos se dividen en g grupos y el total de observaciones se calcula cómo:

$$r\left(\frac{t}{g} + \frac{t}{g} + \frac{t}{g} + \dots + \frac{t}{g}\right) = rg\left(\frac{t}{g}\right) = rt$$

Así, en Gomez y Gomez (1984) no hay $grt= 3(3)(45) = 405$ datos, sino $rt= 3(45) = 135$ observaciones.

Para simplificar cálculos manuales y para homologar ambas metodologías, en algunos denominadores de las fórmulas que se muestran en la sección de resultados, g será considerado nulo, como lo sugirieron González *et al.* (2023) cuando ellos aplicaron submuestreo dentro de parcelas en ensayos monofactoriales en los diseños experimentales completamente al azar, bloques completos al azar y cuadro latino. En esas fórmulas se aplicó la simbología formal descrita en Mendenhall (1987); Sahagún (2007); Montgomery (2009), entre otros.

Software utilizado

Con InfoStat se describe el procedimiento que permitirá la aplicación de la técnica de mínimos cuadrados para la obtención de los análisis de varianza, pero también podrían emplearse InfoGen (<https://www.Info-Gen.com.ar>) o SAS (<https://www.sas.com>), entre otros. Para calcular formas cuadráticas se sugiere usar la calculadora de matrices disponible gratuitamente en <https://www.matrixcalc.org/es/>. Los tres paquetes estadísticos podrían emplearse para generar la comparación de medias de tratamientos dentro de grupos con la prueba de Tukey o diferencia mínima significativa honesta y para su validación aplíquese OPSTAT, <http://14.139.232.166/opstat/default.asp> (Sheoran *et al.*, 1998).

Resultados

Las fórmulas alternativas que generarán las sumas de cuadrados (SC) de los análisis de varianza que fueron publicados en Gomez y Gomez (1984); Shikari *et al.* (2015); Maranna *et al.* (2021), se presentan a continuación:

$$SC \text{ Total} = \sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}^2 - \frac{\left(\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}\right)^2}{rt} = Y'Y - \left(\frac{1}{rt}\right)Y'JY.$$

$$SC \text{ Grupos} = \left(\frac{g}{rt}\right) \sum_{i=1}^g Y_{i..}^2 - \frac{\left(\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}\right)^2}{rt} = \left(\frac{g}{rt}\right)Y'_{i..}Y_{i..} - \left(\frac{1}{rt}\right)Y'JY.$$

$$SC \text{ repeticiones} = \left(\frac{1}{t}\right) \sum_{j=1}^r Y_{.j}^2 - \frac{\left(\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}\right)^2}{rt} = \left(\frac{1}{t}\right)Y'_{.j}Y_{.j} - \left(\frac{1}{rt}\right)Y'JY.$$

En estas fórmulas, $\left(\frac{1}{rt}\right)Y'JY$ es equivalente al factor de corrección utilizado para ajustar las sumas de cuadrados cuando se aplican mínimos cuadrados; Y es un escalar formado por 135 hileras y una columna, Y' es su matriz transpuesta, formada por una hilera y 135 columnas, J es una matriz simétrica formada por 1's, construida con 135 hileras y 135 columnas. Los datos se pueden

tomar de Gomez y Gomez (1984) para verificar la validez de éstas y de las otras fórmulas que se muestran a continuación.

El error a se definirá a partir de: SC trat 1= SC grupos + SC repeticiones + SC error a. Así: SC error a= SC Trat 1- SC grupos- SC repeticiones. Donde:

$$SC\ Trat1 = \left(\frac{g}{t}\right) \sum_{i=1}^g \sum_{j=1}^r Y_{ij}^2 - \frac{\left(\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}\right)^2}{rt} = \left(\frac{g}{t}\right) Y'_{ij} Y_{ij} - \left(\frac{1}{rt}\right) Y' J Y$$

En Gomez y Gomez (1984) el denominador de la primera parte de la formula anterior lo expresaron como $\left(\frac{1}{s}\right)$, equivalente a

$$\left(\frac{1}{s}\right) = \left(\frac{s}{t}\right) = \left(\frac{g}{t}\right), \text{ si } s=g$$

$$SC\ errora = \left(\frac{g}{t}\right) \sum_{i=1}^g \sum_{j=1}^r Y_{ij}^2 - \left(\frac{g}{rt}\right) \sum_{i=1}^g Y_{i..}^2 - \left(\frac{1}{t}\right) \sum_{j=1}^r Y_{.j.}^2 + \frac{\left(\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}\right)^2}{rt}$$

$$= \left(\frac{g}{t}\right) Y'_{ij} Y_{ij} - \left(\frac{g}{rt}\right) Y'_{i.} Y_{i.} - \left(\frac{1}{t}\right) Y'_{.j} Y_{.j} + \left(\frac{1}{rt}\right) Y' J Y$$

Para calcular la primera componente de la formula anterior, deberá construirse una tabla de doble criterio de clasificación: en las hileras se colocarán los grupos, identificados con el subíndice i y en las columnas las repeticiones, representadas con el subíndice j. En ésta habrá $ij=gr=3(3)=9$ datos, lo que implica sumar sobre el subíndice k, correspondiente a cada uno de los subconjuntos de tratamientos que están siendo evaluados; las tres componentes restantes deben calcularse previamente. No debe confundirse el subíndice j, usado para representar a las repeticiones, con la matriz de unos, identificada cómo J; también debe diferenciarse Y, como variable de Y como matriz.

La SC de cada subconjunto de tratamientos en cada grupo se calcula cómo:

$$SC\ TRAT(G1) = \left(\frac{1}{r}\right) \sum_{k=1}^t Y_{1.k}^2 - \left(\frac{g}{rt}\right) \left(\sum_{k=1}^t Y_{1.k}\right)^2 = \left(\frac{1}{r}\right) Y'_{1.k} Y_{1.k} - \left(\frac{g}{rt}\right) Y'_{1.k} J Y_{1.k}$$

$$C\ TRAT(G2) = \left(\frac{1}{r}\right) \sum_{k=1}^t Y_{2.k}^2 - \left(\frac{g}{rt}\right) \left(\sum_{k=1}^t Y_{2.k}\right)^2 = \left(\frac{1}{r}\right) Y'_{2.k} Y_{2.k} - \left(\frac{g}{rt}\right) Y'_{2.k} J Y_{2.k}$$

$$SC\ TRAT(G3) = \left(\frac{1}{r}\right) \sum_{k=1}^t Y_{3.k}^2 - \left(\frac{g}{rt}\right) \left(\sum_{k=1}^t Y_{3.k}\right)^2 = \left(\frac{1}{r}\right) Y'_{3.k} Y_{3.k} - \left(\frac{g}{rt}\right) Y'_{3.k} J Y_{3.k}$$

En éstas se restringe la sumatoria sobre k, y sólo se utilizan 15 datos, tanto en la matriz Y, como en su transpuesta, Y' (15 hileras, una columna, y una hilera, 15 columnas, respectivamente) además, la matriz J está formada por 1's, arreglados en 15 hileras y 15 columnas. El orden de captura es el mismo que en Gomez y Gomez (1984) y lo que se resta es un factor de corrección modificado, generado con la suma de $\left(\frac{t}{g}\right)$ tratamientos en r repeticiones, manteniendo fijo cada uno de los g grupos (G1, G2, G3 o 1, 2, 3).

Sí: $SC_{total} = SC_{grupos} + SC_{repeticiones} + SC_{error\ a} + [SC_{TRAT\ (G1)} + SC_{TRAT\ (G2)} + SC_{TRAT\ (G3)}] + SC_{error\ b}$. Entonces: $SC_{error\ b} = SC_{total} - (SC_{grupos} + SC_{repeticiones} + SC_{error\ a}) - [SC_{TRAT\ (G1)} + SC_{TRAT\ (G2)} + SC_{TRAT\ (G3)}]$. Pero cómo:

$$SC_{Trat1} = SC_{grupos} + SC_{repeticiones} + SC_{error\ a} = \left(\frac{g}{t}\right) Y'_{ij} Y_{ij} - \left(\frac{1}{rt}\right) Y' J Y$$

Entonces, para verificar:

$$SC_{error\ b} = Y' Y - \left(\frac{g}{t}\right) Y'_{ij} Y_{ij} - \left\{ \left[\left(\frac{1}{r}\right) Y'_{1,k} Y_{1,k} - \left(\frac{g}{rt}\right) Y'_{1,k} J Y_{1,k} \right] + \left[\left(\frac{1}{r}\right) Y'_{2,k} Y_{2,k} - \left(\frac{g}{rt}\right) Y'_{2,k} J Y_{2,k} \right] + \left[\left(\frac{1}{r}\right) Y'_{3,k} Y_{3,k} - \left(\frac{g}{rt}\right) Y'_{3,k} J Y_{3,k} \right] \right\}$$

Sí el área experimental se divide en unidad principal (UP) y subunidad (SU) y adicionalmente, se define que la $SC_{total} = SC_{UP} + SC_{SU}$, también se observó que es válida la siguiente expresión: $SC_{UP} = SC_{Trat\ 1} = SC_{grupos} + SC_{repeticiones} + SC_{error\ a}$ así:

$$SC_{UP} = \left(\frac{g}{t}\right) \sum_{i=1}^g \sum_{j=1}^r Y_{ij}^2 - \frac{\left(\sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}\right)^2}{rt} = \left(\frac{g}{t}\right) Y'_{ij} Y_{ij} - \left(\frac{1}{rt}\right) Y' J Y$$

Además, por diferencia: $SC_{SU} = SC_{total} - SC_{UP}$. Donde:

$$SC_{SU} = \sum_{i=1}^g \sum_{j=1}^r \sum_{k=1}^t Y_{ijk}^2 - \left(\frac{g}{t}\right) \sum_{i=1}^g \sum_{j=1}^r Y_{ij}^2 = Y' Y - \left(\frac{g}{t}\right) Y'_{ij} Y_{ij}$$

Uso de InfoStat o InfoGen

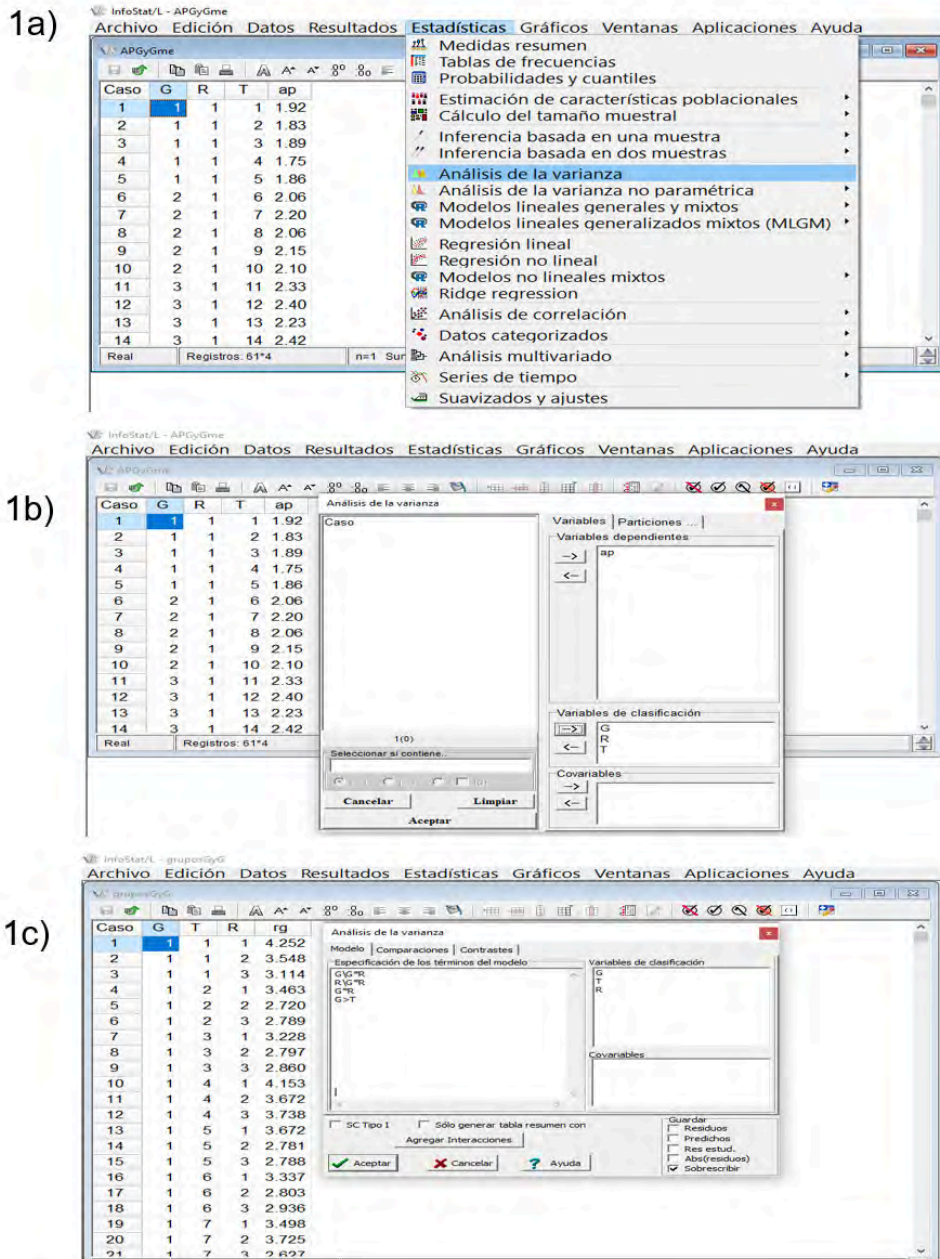
Los rótulos para las columnas serán grupos, repeticiones, tratamientos, y variable respuesta, identificados con G, R, T, Y, respectivamente, pero para evitar confusión con Y, usada para definir la matriz previamente descrita, podría utilizarse otra letra, como X. Los 135 registros serán capturados en tres grupos, cada uno con 15 variedades, en tres repeticiones, en el mismo orden que se muestra en Gomez y Gomez (1984).

Se capturaron algunos datos ficticios de altura de planta (ap; m) de cuatro grupos de variedades de maíz (*Zea mays* L.), sólo para mostrar el procedimiento que deberá aplicarse en este software (Balzarini *et al.*, 2008; Di Rienzo *et al.*, 2008; Balzarini y Di Rienzo, 2016). El análisis estadístico se generará en dos etapas: se mostrará cómo obtener un análisis de varianza general para la partición de efectos en grupos, repeticiones, error a, tratamientos dentro de grupos y error b o residual del modelo.

Las pruebas de F correctas para grupos y repeticiones deben utilizar el cuadrado medio del error a, y variedades dentro de cada grupo el del error b (Figura 1a, b, c), en la segunda etapa se indicará como generar un análisis de varianza individual, considerando la subdivisión de efectos por grupos de tratamientos, por default, repeticiones y tratamientos anidados dentro de grupos se prueban contra el residual del modelo.



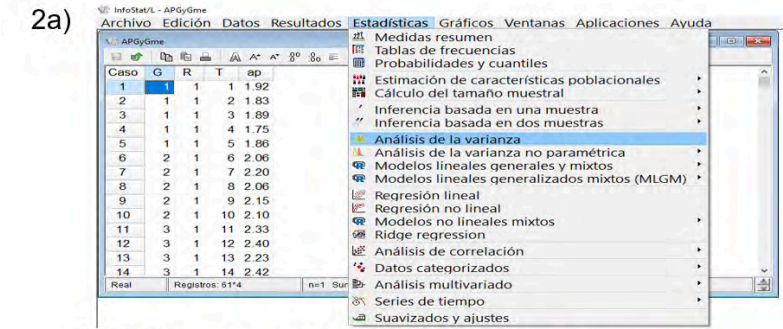
Figura 1. a, b, c. Procedimiento para generar el análisis de varianza general.

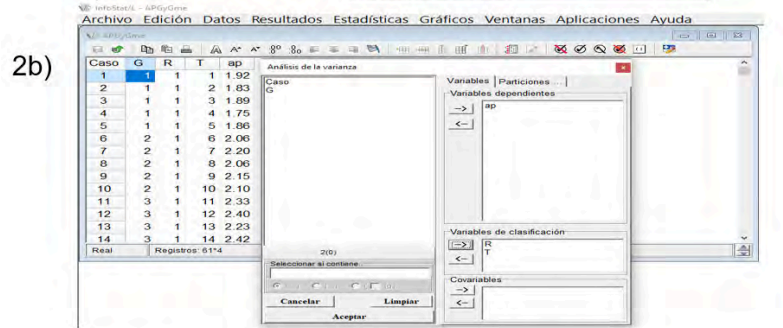


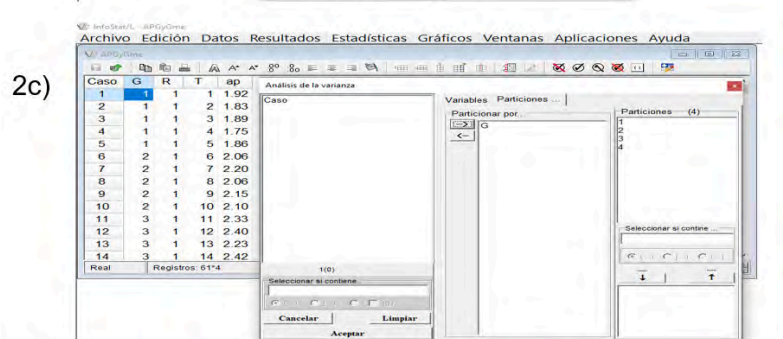
Sin embargo, sus valores de F no son correctos debido a que deben utilizarse los grados de libertad y el cuadrado medio del error b que se muestra en el Anava de la primera etapa. Ambos valores deberán capturarse manualmente en el cuadro de diálogo que mostrará InfoStat (Figura 2 a, b, c, d, e). Así, se observó lo siguiente:

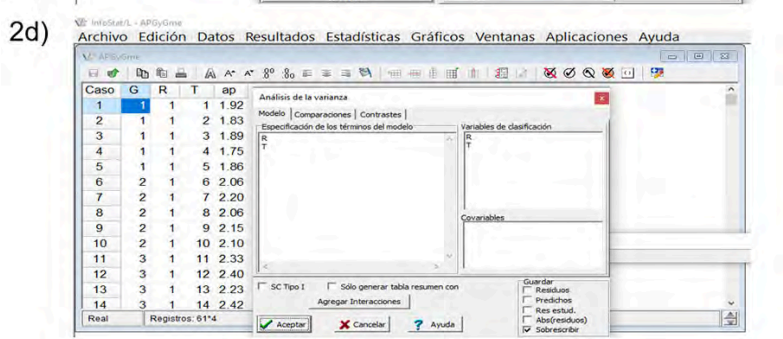


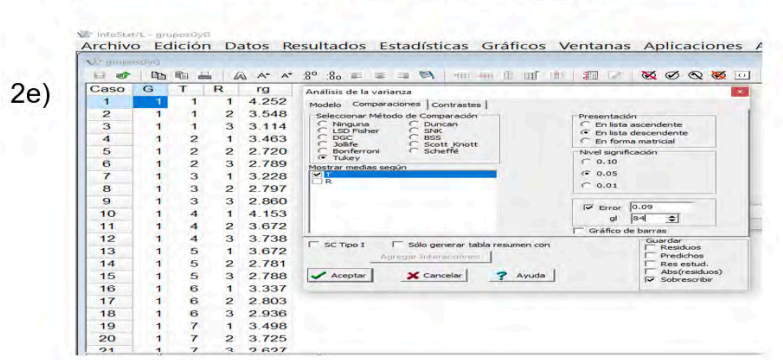
Figura 2 a, b, c, d, e. Subdivisión de los efectos entre tratamientos anidados dentro de grupos.

2a) 

2b) 

2c) 

2d) 

2e) 

Nota: en la imagen anterior deben mostrarse sólo tres grupos, que son los que corresponden a los datos de Gomez y Gomez (1984), se observaron cuatro grupos porque se capturaron datos ficticios en maíz para mostrar el procedimiento en InfoStat.

Discusión

El análisis de grupos contrastantes para subconjuntos de tratamientos con similitudes dentro de cada uno de éstos es de gran relevancia en la investigación agronómica. Esta agrupación podría realizarse considerando diferencias en altura de planta, ciclo biológico, rendimiento de grano u otra característica cuantitativa importante, como lo sugirieron Gomez y Gomez (1984); Shikari *et al.* (2015); Maranna *et al.* (2021), entre otros. También podría clasificarse para los cultivares dentro de cada subconjunto por origen geográfico o genético, como lo hicieron González *et al.* (2008); González *et al.* (2011) pero sin arreglo de bloques completos balanceados.

La importancia de este tipo de estudios también fue destacada por Shikari *et al.* (2015) cuando ellos evaluaron en la India 30 cultivares pertenecientes a tres especies de Brassica anidados en tres grupos de tratamientos, cada uno con 10 materiales genéticamente contrastantes; ellos analizaron el rendimiento de semilla y sus componentes del rendimiento en un arreglo de bloques completos balanceados en un diseño experimental de bloques completos al azar, que en lo sucesivo será denominado como ABCB-DBCA. Para todas las variables se detectaron diferencias altamente significativas entre grupos; entre tratamientos dentro de grupos también las hubo en la mayoría de éstas, representadas en los tipos Gobhi, Yellow y Brown. Los materiales de Gobhi tuvieron el mayor número de días a madurez, pero los de mayor rendimiento de grano por hectárea fueron los del grupo Brown.

Las diversas metodologías estadísticas utilizadas por Maranna *et al.* (2021) en las cuales analizaron integralmente los datos provenientes de un ABCB-DBCA. En la India, ellos evaluaron 68 líneas avanzadas y siete de los mejores cultivares testigo considerando su rendimiento de grano y sus componentes del rendimiento en soya (*Glycine max* (L.) Merrill), su objetivo principal fue identificar material genético de alto rendimiento y mayor estabilidad fenotípica. Se formaron tres grupos de tratamientos con 25 materiales cada uno usando tres repeticiones y considerando días a madurez.

Entre grupos y dentro de éstos se observó amplia variabilidad genética en 12 variables agronómicas. Para la planicie noroeste de la India se concluyó que el material identificado como NRC128 fue estable y superó al mejor testigo (PS 1347) en poco más del 20%. Debido a ésta y otras características favorables que NCR128 tuvo, se hizo su liberación y se recomendó para siembra comercial.

La evaluación de González *et al.* (2008); González *et al.* (2011) de variedades e híbridos de maíz provenientes de diversas empresas y de regiones contrastantes que son recomendables para siembra comercial en el Valle Toluca-Atlaconulco, en el centro del estado de México, México. En estos estudios se aplicó un DBCA sin ABCB para identificar material genético sobresaliente dentro de cada una de las cuatro razas que se localizan en esta región mexicana.

Con la propuesta realizada por Gomez y Gomez (1984); Shikari *et al.* (2015); Maranna *et al.* (2021) se prueban las hipótesis estadísticas para los subconjuntos de tratamientos de manera más precisa en comparación con la realizada por González *et al.* (2008, 2011) u otros investigadores que utilizaron un DBCA sin ABCB debido a que la heterogeneidad ambiental que existe en el área experimental cuando se emplea un ABCB se fracciona en dos componentes: error a y error b. La primera de ambas se utiliza para probar las hipótesis relacionadas con los efectos y varianzas para grupos de tratamientos y para repeticiones, mientras que el error b es utilizado para detectar diferencias significativas entre tratamientos anidados dentro de grupos.

En el contexto anterior, el error a representa la interacción grupos x repeticiones, y el error b es el residual del modelo lineal que se construyó y que fue descrito en el presente estudio. También podría definirse que el error a está relacionado con la unidad principal y que el error b está asociado con la subunidad. En González *et al.* (2008, 2010, 2011) o en múltiples ensayos donde se evalúen

ensayos de rendimiento para evaluar los efectos entre tratamientos con otra opción, como sin o con el uso de contrastes mutuamente ortogonales, éstos se prueban con el residual del modelo, que es equivalente a su error experimental, las pruebas de F para tratamientos dependen de si existen o no diferencias estadísticas entre grupos de materiales en un ABCB.

Las notaciones sumatoria y punto han sido de gran utilidad para representar y analizar datos de variables cuantitativas en múltiples ensayos agronómicos, pero como lo han mostrado Gomez y Gomez (1984); Martínez (1988); Cochran y Cox (2004), entre otros, éstas podrían expresarse algebraica o matricialmente usando cualquier simbología cuando se realizan cálculos manuales. En contraposición Mendenhall (1987); Zamudio y Alvarado (1996); Sahagún (1998); Montgonery (2009); Restrepo (2007a, 2007b), entre otros, utilizan la simbología formal para evitar confusión en la descripción de dichos procedimientos, particularmente, al describir las guías para la construcción de modelos fijos, aleatorios, o mixtos o cuando se aplican las reglas para obtener las varianzas de interés a partir de la esperanza matemática de los cuadrados medios.

Además de la simbología utilizada, otros aspectos que causan confusión durante los cálculos o en el manejo de un paquete estadístico es la ausencia del modelo lineal que fue aplicado, así como el tipo de efectos que están siendo evaluados, especialmente en experimentos factoriales complejos; los cálculos manuales frecuentemente se consideran como un prerrequisito para la aplicación de software.

En el contexto anterior, en el presente estudio se homologó el modelo estadístico con la aplicación de dos metodologías para calcular grados de libertad y sumas de cuadrados como un prerrequisito para lograr lo mencionado previamente, InfoGen, InfoStat o SAS, entre otros, serán de gran utilidad para lograr esta meta.

Si el área experimental, como fue presentada por Gomez y Gomez (1984), correspondiente a un ABCB-DBCA, es dividida en unidad principal (UP) y subunidad (SU), sus grados de libertad se calcularían como $rg-1$ y $r(t-g)$, respectivamente, cuya suma origina los $rt-1$ grados de libertad del total, tanto en este arreglo como en un DBCA sin ABCB. Adicionalmente, el total para grados de libertad de tratamientos dentro de grupos sería:

$$\sum_{i=1}^g (t_i - 1)$$

que es equivalente a $t - g$. Así, también sería más fácil calcular los grados de libertad para el error b como $(r-1)(t-g)$, siendo r , t , g , los niveles para repeticiones, tratamientos y grupos, respectivamente.

Los resultados de González *et al.* (2019) también fraccionaron los efectos entre tratamientos en grupos de materiales genéticos evaluados en un DBCA pero sin ABCB, dichos efectos fueron analizados por medio de la técnica de contrastes mutuamente ortogonales pero la precisión con la que se prueban las hipótesis estadísticas de interés son más confiables cuando se utiliza un ABCB-DBCA, debido a que se usan los errores a y b, aunque éstos últimos también dependen de la magnitud de sus grados de libertad y de la existencia de diferencias significativas entre subconjuntos de tratamientos.

En un DBCA sin ABCB, la suma de cuadrados (SC) de tratamientos es igual a SC de grupos + SC de tratamientos dentro de grupos. También, la SC del error experimental es igual a la adición de las SC de los errores a y b que se obtienen cuando se genera el Anava combinando los datos de los g grupos. Finalmente, la SC de repeticiones es igual a la diferencia que se origina entre la SC de repeticiones dentro de cada grupo y la SC del error a. En el contexto anterior, como se indicó anteriormente, el error a representa la interacción entre grupos y repeticiones, asociado a la unidad principal y el error b es el residual del modelo estadístico (ϵ_{ijk}) que se mostró en la sección de materiales y métodos.

Debido a que deben ser verificados varios cálculos manuales antes de aplicarse algún paquete estadístico, deberá checarsse que la adición de la SC del error en cada uno de los g grupos más el correspondiente a la SC del error a sea igual a la SC del error experimental en la salida que se origine aplicando un DBCA sin ABCB. Con relación a lo anterior, las salidas que origine InfoStat, InfoGen o SAS, entre otros, serán de gran utilidad para verificar los cálculos que serán originados cuando se utilicen las fórmulas previamente presentadas.

Autores como Mendenhall (1987); Sahagún (1998); Montgomery (2009) puntualizaron el hecho de que el análisis de varianza es una parte importante para enfrentar el problema que representa el diseño y análisis de cualquier ensayo experimental, en el cual está involucrado el cálculo de grados de libertad, sumas de cuadrados y la construcción de pruebas estadísticas apropiadas considerando la relación que existe entre los cuadrados medios y sus esperanzas matemáticas, especialmente cuando se consideran modelos aleatorios o mixtos en situaciones más complejas. Esta problemática también ha sido destacada por otros investigadores como Montgomery (2009); Restrepo (2007a, 2007b); Piepho *et al.* (2003).

La investigación de González *et al.* (2023) hicieron énfasis en introducir correctamente las instrucciones o los procedimientos en especificación en los términos del modelo en InfoStat o InfoGen o en el editor de SAS, para probar adecuadamente las hipótesis estadísticas relacionadas con los experimentos conducidos, sin y con submuestreo dentro de las unidades experimentales, cuando se aplican los diseños experimentales completamente al azar, bloques completos al azar y cuadro latino, Zamudio y Alvarado (1996) hicieron la misma recomendación cuando elaboraron diversos códigos para SAS, para analizar los tres diseños experimentales previamente mencionados en submuestreo balanceado.

En el presente estudio, como lo hicieron Gomez y Gomez (1984); Shikari *et al.* (2015); Maranna *et al.* (2021), el error a debe utilizarse para probar los efectos entre grupos y entre repeticiones, y el error b es empleado para realizar lo mismo con los subconjuntos de tratamientos anidados dentro de grupos (imágenes 4 a 8 de resultados).

Para la comparación de medias de variedades dentro de grupos, InfoStat e InfoGen son muy flexibles, debido a que en ambos la base de datos se ordena automáticamente (última imagen de resultados) y adicionalmente, permiten realizar la corrección que debe efectuarse a la diferencia mínima significativa honesta o prueba de Tukey, realizando la captura manual de los grados de libertad y del cuadrado medio del error b, generados con todos los datos registrados en el experimento. Si las diferencias entre grupos de tratamientos no son significativas, InfoStat puede generar un análisis de varianza y una comparación de medias con las Prueba de Tukey utilizando la misma base de datos que cuando se usa un ABCA-DBCA. Su validación podría efectuarse con el software Opstat, disponible gratuitamente en su sitio WEB, en el cual sólo se capturan las medias aritméticas de cada variedad dentro de cada grupo, así como los grados de libertad y el cuadrado medio del error b, los cuales pueden generarse con cualquier software o, más fácilmente, con la hoja de cálculo de Microsoft Excel.

Conclusiones

El modelo estadístico apropiado se construyó considerando que los tratamientos están anidados dentro de grupos y que éstos últimos están cruzados con repeticiones. El cálculo manual de grados de libertad y sumas de cuadrados se simplificó al dividir el área experimental en unidad principal y subunidad, en ambos se alojan los errores a y b, respectivamente, la primera representa la interacción grupos x repeticiones y la segunda es el residual del modelo estadístico bajo consideración.

Las técnicas de mínimos cuadrados y formas cuadráticas o matriciales generan resultados similares pero la primera es más fácil de utilizar cuando se aplica un paquete estadístico, especialmente si el número de variables por analizar tiende a aumentar. InfoStat es muy flexible cuando se aplica la prueba de Tukey a los tratamientos anidados dentro de grupos, debido a que permite corregir la diferencia mínima significativa honesta cuando se capturan manualmente los grados de libertad y el cuadrado medio del error b. Si los grupos de tratamientos son iguales estadísticamente, el ensayo puede analizarse como un DBCA utilizando la misma base de datos que un ABCB-DBCA, como opción para generar los mismos resultados puede recurrirse al software Opstat, disponible gratuitamente en su sitio WEB.

Bibliografía

- 1 Balzarini, M. G. y Di Rienzo, J. A. 2016. InfoGen. FCA. Universidad Nacional de Córdoba, Argentina. <http://www.info-Gen.com.ar>.
- 2 Balzarini, M. G.; González, L.; Tablada, M.; Casanoves, F.; Di Rienzo, J. A. y Robledo, C. W. 2008. Manual del usuario de InfoStat. Ed. Brujas, Córdoba, Argentina. 348 p.
- 3 Cochran, W. G. y Cox, G. M. 2004. Diseños experimentales. Ed. Trillas SA de CV, 6^{ta}. Reimpresión. México, DF. 661 p.
- 4 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. (<https://www.infostat.com.ar>).
- 5 Gomez, K. A. y Gomez, A. A. 1984. Statistical procedures for agricultural research. 2nd Ed. John Wiley & Sons, Inc. Printed in Singapore. 680 p.
- 6 González, H. A.; Vázquez, G. L. M.; Sahagún, C. J. y Rodríguez, P. J. E. 2008. Diversidad fenotípica de variedades e híbridos de maíz en el Valle Toluca-Atlaconulco, México. Revista Fitotecnia Mexicana. 31(1):67-76.
- 7 González, H. A.; Pérez, D. J.; Sahagún, C. J.; Franco, M. O.; Morales, E. J.; Rubí, A. M.; Gutiérrez, R. F.; Balbuena, A. 2010. Aplicación y comparación de métodos univariados para evaluar la estabilidad en maíces del Valle Toluca-Atlaconulco, México. Revista Agronomía Costarricense. 34(2):129-143.
- 8 González, H. A.; Pérez, L. D. J.; Franco, M. O.; Nava, B. E. B.; Gutiérrez, R. F.; Rubí, A. M. y Castañeda, V. A. 2011. Análisis multivariado aplicado al estudio de las interrelaciones entre cultivares de maíz y variables agronómicas. Revista Ciencias Agrícolas Informa. 20(2):58-65.
- 9 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. Revista Mexicana de Ciencias Agrícolas. 10(6):1417-1431.
- 10 González, H. A.; Pérez, L. D. J.; Balbuena, M. A.; Franco, M. J. R.; Gutiérrez, R. F. y Rodríguez, G. J. A. 2023. Submuestreo balanceado en experimentos monofactoriales usando InfoStat y InfoGen : validación con SAS. Revista Mexicana de Ciencias Agrícolas . 14(2):235-249.
- 11 Little, T. M. y Hills, F. J. 2008. Métodos estadísticos para la Investigación en la agricultura. Ed. Trillas, SA. de CV. México, DF. 270 p. ISBN:978-968-24-3629-1.
- 12 Maranna, S.; Nataraj, V.; Kumawat, G.; Chandra, S.; Rajesh, V.; Ramteko, R.; Manohar Patel, R.; Ratnaparkhe, M. B.; Husain, S. M.; Gupta, S. and Khandekar, N. 2021. Breeding for higher yield, early maturity, wider adaptability and wáterlogging tolerance in soybean (*Glycine max* L.): a case study. Scientific Reports. 11:22853-<https://doi.org/10.1038/s41598-021-02064-x>.
- 13 Martínez, G. A. 1988. Diseños experimentales. métodos y elementos de teoría . Editorial. Trillas, 1^{ra} Ed. México, DF. 756 p.
- 14 Mendenhall, W. 1987. Introducción a la probabilidad y la estadística. Grupo Editorial Iberoamérica. 1^{ra} Ed. México, DF. 626 p.
- 15 Montgomery, D. C. 2009. Design and analysis of experiments. 7th Ed. John Wiley (Sons, Inc. USA. 656 p.
- 16 Piepho, H. P.; Büsche, A. and Enrich, K. 2003. A Hitchhiker's guide to mixed models for randomized experiments. J. Agron. Crop Sci. 189(2):310-322.

- 17 Restrepo, L. F. 2007 a. Diagramas de estructuras en el análisis de varianza. Revista Colombiana de Ciencias Pecuarias. 20(2):202-208.
- 18 Restrepo, B. L. F. 2007 b. La esperanza del cuadrado medio. Revista Colombiana de Ciencias Pecuarias . 20(2):193-201.
- 19 Sahagún, C. J. 1998. Construcción y análisis de los modelos fijos, aleatorios y mixtos. Departamento de Fitotecnia. Programa Nacional de Investigación en Olericultura. Universidad Autónoma Chapingo (UACH). Boletín técnico núm. 2, 64 p.
- 20 Sahagún, C. J. 2007. Estadística descriptiva y probabilidad: Una perspectiva biológica. 2^{da}. Ed. Universidad Autónoma Chapingo (UACH). México, DF. 282 pp. ISBN: 978 -968-02-0357-4.
- 21 Sheoran, O. P.; Tonk, D. S.; Kaushik, L. S.; Hasija, R. C. and Pannu, R. S. 1998. Statistical Software package for Agricultural research workers. Recent Advances in Information Theory, Statistical (Computer Applications by D.S. Hooda (R. C. Hasija Department of Mathematics Statistics, CCS HAU, Hisar. 139-143 pp.
- 22 Shikari, A. B.; Pourray, G. A.; Sofi, N. R.; Hussain, A.; Dar, Z. A. and Iqbal, A. M. 2015. Group balanced block design for comparisons among oilseed Brassicae. Academic Journals. 10(8):302-305. <https://doi.org/10.5897/SRE2014.5792>.
- 23 Zamudio, S. F. J. y Alvarado, S. A. A. 1996. Análisis de diseños experimentales con igual número de submuestras. División de Ciencias Forestales. Universidad Autónoma Chapingo (UACH). Primera Edición. México, DF. 85 p. ISBN: 968 884 489 6.



Tratamientos anidados dentro de un arreglo en grupos de bloques completos balanceados

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 January 2024
Date accepted: 01 February 2024
Publication date: 14 March 2024
Publication date: March 2024
Volume: 15
Issue: 2
Electronic Location Identifier: e3634
DOI: 10.29312/remexca.v15i2.3634

Categories

Subject: Artículo

Palabras clave:

Palabras clave:

diseño de bloques completos al azar

formas cuadráticas

InfoStat

mínimos cuadrados

Counts

Figures: 2

Tables: 0

Equations: 16

References: 23

Pages: 0