Revista Mexicana de Ciencias Agrícolas

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Artículo

Serie de experimentos para tratamientos anidados en grupos con arreglo de bloques completos balanceados

Andrés González-Huerta[1,§]
Delfina de Jesús Pérez-López[1]
Jesús Hernández-Ávila[1]
J. Ramón Pascual Franco-Martínez[1]
Artemio Balbuena-Melgarejo[1]
Martín Rubí-Arriaga[1]

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

Autor para correspondencia: agonzalezh@uaemex.mx

Resumen

El diseño y el análisis de diversas series de experimentos a través de años, localidades o ambos ha sido de gran relevancia en la investigación agronómica. En este estudio se extrapola, a través de ambientes, el caso presentado por Gomez y Gomez (1984) con relación al rendimiento de grano registrado en 45 variedades de arroz anidadas en tres grupos, en arreglo de bloques completos balanceados; se construye su modelo estadístico para un diseño experimental de bloques completos al azar, se incluyen las fórmulas para calcular sumas de cuadrados y se propone el procedimiento para generar una salida si fuera aplicado InfoGen. Los procedimientos se basan en: ambientes, grupos, ambientes x grupos, repeticiones dentro de ambientes, error a, ambientes x tratamientos dentro de grupos, tratamientos dentro de grupos y error b; las primeras cinco componentes definen la unidad principal (UP), y las restantes son la subunidad (SUB). Adicionalmente, son mencionadas otras formas para calcular grados de libertad para UP y SUB, así como los correspondientes al residual del modelo o error b, las cuales simplifican los cálculos manuales. 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 y, finalmente, se indica cómo aplicar la prueba de Tukey para la comparación de medias de variedades dentro de cada grupo si es empleado este software, InfoStat o SAS.

Palabras clave:

diseño de bloques completos al azar, InfoStat, suma de cuadrados y formulas cuadráticas.

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i7.3831

Introducción

El diseño, el análisis y la interpretación de los datos provenientes de un experimento, o de una serie de éstos registrados en años, localidades o en ambos, se ha convertido en una herramienta muy importante en las ciencias agropecuarias cuando se usa un diseño experimental completamente al azar, bloques completos al azar, cuadro latino o algún tipo de látice, aunque han sido más empleados los dos primeros (Sahagún, 1993; Martínez, 1988; Sahagún, 2007).

En un DBCA convencional, en cada ensayo los tratamientos son asignados al azar de manera independiente en cada repetición, de tal forma que todos los bloques sean perpendiculares al gradiente de variabilidad predominante, como pendiente o nivel de fertilidad en un suelo. Generalmente, los bloques son del mismo tamaño y cada tratamiento es asignado una sola vez dentro de cada uno de éstos. Siempre que sea posible, las diferencias dentro de cada bloque deberán ser muy pequeñas y la heterogeneidad entre éstos debe tender a maximizarse para que este diseño experimental sea eficiente (Gomez y Gomez, 1984; Martínez, 1988; Montgomery, 2009; Little y Hills, 2008).

En áreas experimentales homogéneas no es justificable el uso de un DBCA; con dos gradientes de variabilidad, uno perpendicular al otro, tendría que elegirse un cuadro latino o un látice, dependiendo de qué tan grande sea el número de tratamiento y de repeticiones que serán evaluadas; en los látices parcialmente balanceados frecuentemente se evalúan más de 30 tratamientos (Cochran y Cox, 1954; Gomez y Gomez, 1984; Martínez, 1998).

En el contexto anterior, para un DBCA habría dos posibilidades: sin y con arreglo de bloques completos balanceados. En Gomez y Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021) se abordan los conceptos y la aplicación de la segunda posibilidad: cada repetición es subdividida en grupos de tratamientos que comparten alguna similitud dentro de éstos y que difieren de manera importante entre ellos.

En la formación de dichos grupos podría considerarse el ciclo biológico de los cultivares o su altura de planta, pero esta agrupación también podría hacerse considerando su origen genético y geográfico como lo han sugerido González *et al*. (2008, 2010, 2011), entre otros. Según González *et al*. (2024), para un solo ensayo donde se utilice un diseño experimental de bloques completos al azar en arreglo de bloques completos balanceados (DBCA-ABCB), el área experimental podría dividirse en unidad principal (UP) y subunidad (SUB), de tal forma que en la primera estén alojados los grupos (G), las repeticiones (R), y el error a, este último equivalente a la interacción GxR, mientras que la segunda deberá contener a tratamientos anidados dentro de G, [T(G)], y al error b.

Debido a la estrecha relación que existe entre un experimento monofactorial y las factoriales que usan el mismo diseño experimental, este enfoque también podría ser validado para series de experimentos en DBCA-ABCB, como puede inferirse de González *et al*. (2024). En el marco de referencia previo es necesario extrapolar este tipo de estudios al caso de años, localidades o ambos en una serie de experimentos a través de ambientes, como un prerrequisito para evaluar nuevo material en un programa de mejoramiento genético, así como para validar, aplicar, generar o transferir tecnología a campos de productores. Así, el objetivo principal de esta investigación fue generar el modelo estadístico y las fórmulas para calcular grados de libertad y sumas de cuadrados para un arreglo de unidades experimentales, como el que es mencionado previamente.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

Materiales y métodos

Conceptos preliminares

El fitomejorador colecta, evalúa e identifica material biológico sobresaliente en un programa de mejoramiento vegetal aplicando algún diseño genético y experimental adecuado a través de varias localidades y algunos años para que los resultados sean más confiables. Esta situación deberá representarse realísticamente en un modelo estadístico el cual debe ser independiente de la disponibilidad de infraestructura de equipo electrónico; es el fitomejorador y su equipo asesor los responsables de su correcta construcción.

En este deberá determinarse si los factores son fijos o aleatorios y si entre éstos existe cruzamiento o anidamiento. También deberá considerarse la relación que existe entre los componentes del modelo y un análisis de varianza y una comparación de medias de tratamientos, por medio de la estimación de efectos o varianzas (Sahagún, 1998). Antes de definir los tipos de modelos que frecuentemente se usan en las diferentes ramas de las ciencias agropecuarias considere el caso de la raza de maíz Cacahuacintle.

Una población de variedades criollas de esta raza podría estar determinada por el número de agricultores que la siembran en el municipio de Calimaya de Díaz González, en el estado de México. Si en este municipio hay 2500 agricultores entonces habrá 2500 criollos, suponiendo que cada uno de estos disponga de una variedad diferente. Un modelo es de efectos fijos si el investigador considera una muestra aleatoria y representativa de las 2500 variedades; por ejemplo, de tamaño 30 y estima las diferencias que existen sólo entre éstas, pero si a partir de ellas hace inferencia hacia toda la población de criollos entonces estará eligiendo un modelo de efectos aleatorios; en el primer caso se estiman efectos y en el segundo varianzas.

Los modelos estadísticos se construyen usando ambos principios y éstos pueden ser fijos, aleatorios o mixtos; esta última situación surge de la necesidad de incluir componentes tanto fijas como aleatorias. Como ejemplo considere la necesidad de evaluar 30 variedades de Cacahuacintle en tres localidades del Valle de Toluca, en el estado de México, usando tres repeticiones por tratamiento. Si variedades y localidades son considerados factores fijos y aleatorios, respectivamente, entonces se tendrá un modelo mixto.

En la investigación agronómica es muy frecuente el uso de modelos cuyas componentes son fijas y aleatorias, como en las series de experimentos a través del tiempo, del espacio o en ambos que fueron discutidas en Sahagún (1993, 2007) o como la que fue considerada en Gomez y Gomez (1984); Maranna *et al*. (2021); González *et al*. (2024). Los años, las localidades o las combinaciones entre éstos generan componentes aleatorias en sus modelos estadísticos (Sahagún, 1998).

En el presente estudio se dice que los factores A y G, usados para identificar ambientes y grupos, respectivamente, están cruzados, cuando cada nivel de A está combinado con cada nivel de G. El factor T, que representa variedades o tratamientos, está anidado en el factor G, si cada nivel de T se combina con sólo un nivel de G. En las series de experimentos a través de años, localidades o ambos, las repeticiones también están anidadas dentro de éstos.

Adicionalmente, las parcelas o unidades experimentales están anidadas en repeticiones y en localidades. Si un factor está anidado en otro no es posible estudiar su interacción. Un Conjunto de datos es balanceado si el número de observaciones en cada celda de menor tamaño que se puede formar es una constante (Sahagún, 1998). En los diseños experimentales completamente al azar, bloques completos al azar y cuadro latino se tienen experimentos balanceados cuando cada tratamiento tiene el mismo número de repeticiones y cuando en cada parcela o unidad experimental se registra el mismo número de observaciones; esta última situación comúnmente está relacionada con submuestreo en diseños experimentales.

De otra manera se tendrá una situación desbalanceada; sin balance, el análisis estadístico de los datos es más complejo. La esperanza del cuadrado medio es imprescindible cuando se desea racionalizar la metodología usada para probar una hipótesis en un análisis de varianza o para

DOI: https://doi.org/10.29312/remexca.v15i7.3831

estimar componentes de varianza; éstas pueden derivarse utilizando resultados del modelo lineal general o generarse directamente. En este contexto, existen diversas publicaciones en las cuales se proporcionan guías para la construcción de modelos estadísticos o esperanzas del cuadrado medio, como en Sahagún (1998); Restrepo (2007 a, b); Piepho *et al*. (2003), entre otros.

Modelo estadístico

El modelo para una serie de experimentos para evaluar tratamientos anidados dentro de grupos en arreglo de bloques completos balanceados, en un diseño experimental de bloques completos al azar, se construyó con base en la guía proporcionada por Sahagún (1998); Piepho *et al*. (2003); Restrepo (2007a). Este es: $X_{ijkl}= \mu+ A_i + G_j + R_{k(i)} + (AG)_{ij} + (GR)_{jk(i)} + \tau_{l(j)} + (AT)_{il(j)} + \varepsilon_{ijkl}$.

Dónde: X es el rendimiento de grano o cualquier otra variable cuantitativa, µ es la media aritmética de los art datos, A_i es el efecto causado por el i-ésimo ambiente, G_j es el efecto causado por el j-ésimo grupo, $R_{k(i)}$ es la contribución de la k-ésima repetición anidada en el i-ésimo ambiente, $(AG)_{ij}$ es la interacción que originan los niveles ij de los factores A y G, $(GR)_{jk(i)}$ es la interacción del j-ésimo grupo con la k-ésima repetición anidada en el i-ésimo ambiente, también denominada error a, $\#_{l(j)}$ es el efecto causado por el l-ésimo tratamiento anidado dentro del j-ésimo grupo, $(AT)_{il(j)}$ es la interacción entre los niveles de los factores A y T, éste último anidado dentro de grupos, y $\#_{ijkl}$ es el residual del modelo, también conocido como error b.

Simbología utilizada para calcular suma de cuadrados

Las variables de clasificación en el modelo construido previamente son ambientes, grupos, repeticiones y tratamientos, las cuales han sido identificadas con los subíndices i, j, k, l; sus niveles son a, g, r, t/g, respectivamente. En el presente estudio siemre g = 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:

$$ar\left(\frac{t}{g}+\frac{t}{g}+\frac{t}{g}+,\ldots,+\frac{t}{g}\right)= arg\left(\frac{t}{g}\right) = art.$$

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 Sahagún (2007); Montgomery (2009); Mendenhall (1987), entre otros.

Software utilizado

Con InfoGen 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 InfoStat (https://www.InfoStat.com.ar) o SAS (https://www.sas.com), entre otros. Los tres paquetes estadísticos podrían utilizarse 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 también podría aplicarse OPSTAT (http://14.139.232.166/opstat/default.asp; Sheoran *et al*., 1998).

Resultados

Fórmulas para calcular grados de libertad (GL y suma de cuadrados (SC

Las fórmulas que generarán GL y SC en los análisis de varianza de una serie de experimentos a través de ambientes, para el tipo de arreglo de unidades experimentales mencionado previamente, se presentan a continuación y, éstas son una extensión de las que fueron publicadas para un experimento mono factorial por González et al. (2024).

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i7.3831

Fórmulas para calcular GL

GL total= art -1. GL ambientes (A)= a - 1. GL grupos (G) = g - 1. GL repeticiones dentro A= a(r-1). GL A x G= (a -1) (g -1). GL error a= a(g-1) (r-1). GL tratamientos (T) anidados en G= t - g. GL AxT(G)= (a-1) (t-g). GL error b= a(r-1) (t-g). Si el área experimental es dividida en unidad principal (UP) y subunidad (SU) (González *et al*., 2024), sus GL serían, respectivamente, agr -1 y ar(t-g). La suma de ambos es art -1, que corresponde a GL total.

Fórmulas para calcular SC

SC total=

$$\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{eart} = Y'Y - \left(\frac{1}{art}\right)Y'JY$$

SC ambientes (A)=

$$\left(\frac{1}{rt}\right)\sum_{i=1}^{a} Y_{i...}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{art} = \left(\frac{1}{rt}\right)Y'_{i...}Y_{i...} - \left(\frac{1}{art}\right)Y'JY$$

SC grupos=

$$\left(\frac{g}{art}\right)\sum_{j=1}^{g} Y_{.j..}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{art} = \left(\frac{g}{art}\right)Y'_{.j..}Y_{.j..} - \left(\frac{1}{art}\right)Y'JY$$

SC AxG=

$$\left\{\left(\frac{1}{rt}\right)\sum_{i=1}^{a}\sum_{j=1}^{g} Y_{ij..}^2 - \left(\frac{\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}}{art}\right)^2\right\} - \text{SC A} - \text{SC G.}$$

SC R(A)=

$$\left\{\left(\frac{1}{t}\right)\sum_{i=1}^{a}\sum_{k=1}^{r} Y_{i.k.}^2 - \left(\frac{\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}}{art}\right)^2\right\} - \text{SC A.}$$

SC error a =

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

El error a está contenido en la fórmula:

SC trat 1=

$$\left(\frac{g}{t}\right)\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r} Y_{ijk.}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{art} = \left(\frac{g}{t}\right)Y'_{ijk.}Y_{ijk.} - \left(\frac{1}{art}\right)Y'JY$$

También:

SC trat 1= SC A + SC G + SC AxG + SC R(A) + SC error a.

Por lo tanto: SC error a= SC trat 1 - SC A - SC G - SC AxG - SC R(A). En la ecuación anterior obsérvese que SC UP= SC trat 1.

Para verificación: SC error a:

$$\left(\frac{g}{t}\right)\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r} Y_{ijk.}^2 - \left(\frac{1}{rt}\right)Y'_{i...}Y_{i...} - \left(\frac{g}{art}\right)Y'_{.j..}Y_{.j..} - \left(\frac{1}{at}\right)Y'_{..k.}Y_{..k.} - \left(\frac{g}{rt}\right)Y'_{ij..}Y_{ij..} + \left(\frac{3}{art}\right)Y'JY$$

En las fórmulas previas,

DOI: https://doi.org/10.29312/remexca.v15i7.3831

$(\left(\frac{1}{art}\right))\ Y'JY$

es equivalente al factor de corrección utilizado para ajustar las sumas de cuadrados; Y será un escalar formado por 270 hileras y una columna, Y' su matriz transpuesta, estará formada por una hilera y 270 columnas, J será una matriz simétrica formada por 1's, construida con 270 hileras y 270 columnas.

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 j, y en las columnas los ambientes y las repeticiones, representadas con los subíndices i, k, respectivamente. En ésta habrá ijk= agr= 2(3)(3)= 18 datos, lo que implica sumar sobre el subíndice l, correspondiente a cada uno de los subconjuntos de tratamientos que están siendo evaluados; las cinco componentes restantes deberán calcularse previamente.

No debe confundirse el subíndice j, usado para representar a los grupos, con la matriz de unos, identificada cómo J; también debe diferenciarse Y, como variable de Y como matriz. La SC de tratamientos anidados dentro de grupos se calculará cómo:

SC TRAT (G1)=

$$\left(\frac{1}{ar}\right)\sum_{l=1}^{t} Y^2_{.1.l} - \left(\frac{g}{art}\right)\left(\sum_{l=1}^{t} Y_{.1.l}\right)^2 = \left(\frac{1}{ar}\right)Y'_{.1.l}Y_{.1.l} - \left(\frac{g}{art}\right)Y'_{.1.l}J\ Y_{.1.l}$$

SC TRAT (G2)=

$$\left(\frac{1}{ar}\right)\sum_{l=1}^{t} Y^2_{.2.l} - \left(\frac{g}{art}\right)\left(\sum_{l=1}^{t} Y_{.2.l}\right)^2 = \left(\frac{1}{ar}\right)Y'_{.2.l}Y_{.2.l} - \left(\frac{g}{art}\right)Y'_{.2.l}J\ Y_{.2.l}$$

.

SC TRAT (G3)=

$$\left(\frac{1}{ar}\right)\sum_{l=1}^{t} Y^2_{.3.l} - \left(\frac{g}{art}\right)\left(\sum_{l=1}^{t} Y_{.3.l}\right)^2 = \left(\frac{1}{ar}\right)Y'_{.3.l}Y_{.3.l} - \left(\frac{g}{art}\right)Y'_{.3.l}J\ Y_{.3.l}$$

La suma de cuadrados del g-ésimo grupo se calculará similarmente:

SC TRAT (Gg)=

$$\left(\frac{1}{ar}\right)\sum_{l=1}^{t} Y^2_{.g.l} - \left(\frac{g}{art}\right)\left(\sum_{l=1}^{t} Y_{.g.l}\right)^2 = \left(\frac{1}{ar}\right)Y'_{.g.l}Y_{.g.l} - \left(\frac{g}{art}\right)Y'_{.g.l}J\ Y_{.g.l}$$

Para verificar que los cálculos previos sean correctos, el total de todas las SC de tratamientos dentro de grupos será:

SC T(G):

$$\left(\frac{1}{ar}\right)\sum_{l=1}^{t} Y^2_{...l} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{art} = \left(\frac{1}{ar}\right)Y'_{...l}Y_{...l} - \left(\frac{1}{art}\right)Y'JY$$

Para definir la suma de cuadrados de la interacción entre los ambientes y los tratamientos anidados dentro de grupos, primero establézcase la siguiente relación:

SC trat 5= SC A + SC G + SC AxG + SC T(G) + SC AxT(G). Dónde:

SC trat 5=

$$\left(\frac{1}{r}\right)\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{l=1}^{t} Y^2_{ij.l} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t} Y_{ijkl}\right)^2}{art} = \left(\frac{1}{r}\right)Y'_{ij.l}Y_{ij.l} - \left(\frac{1}{art}\right)Y'JY$$

En este contexto:

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i7.3831

SC A x T(G) = SC trat 5 - SC A - SC G - SC AxG - SC T(G).=

$$\left(\frac{1}{r}\right)Y'_{ij.l}Y_{ij.l} - \left(\frac{1}{rt}\right)Y'_{i...}Y_{i...} - \left(\frac{g}{art}\right)Y'_{.j..}Y_{.j..} - \left(\frac{g}{rt}\right)Y'_{ij..}Y_{ij..} - \left(\frac{1}{ar}\right)Y'_{...l}Y_{...l} + \left(\frac{3}{art}\right)Y'JY$$

Adicionalmente:

SC total= SC A + SC G + SC R(A) + SC AxG + SC error a + [SC TRAT (G1) + SC TRAT (G2) + SC TRAT (G3) + ,…, + SC TRAT (Gg)] + SC AxT(G) + SC error b.

Entonces:

SC error b= SC total - (SC A + SC G + SC R(A) + SC AxG + SC error a) - [SC TRAT (G1) + SC TRAT (G2) + SC TRAT (G3) + ,…, + SC TRAT (Gg)] - SC AxT(G).

Sí el área experimental se divide en unidad principal (UP) y subunidad (SU) y como lo propusieron González *et al*. (2024), se define que la SC total= SC UP + SC SU, entonces también será válida la siguiente expresión: SC UP= SC A + SC G + SC R(A) + SC AxG + SC error a. En el contexto anterior, también es correcta la siguiente equivalencia:

SC UP= SC trat 1=

$$\left(\frac{g}{t}\right)\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}Y_{ijk.}^{2} - \frac{\left(\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t}Y_{ijkl}\right)^{2}}{art} = \left(\frac{g}{t}\right)Y'_{ijk.}Y_{ijk.} - \left(\frac{1}{art}\right)Y'JY$$

Por diferencia: SC SU= SC total - SC UP. Por lo tanto:

SC SU=

$$\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}\sum_{l=1}^{t}Y_{ijkl}^{2} - \left(\frac{g}{t}\right)\sum_{i=1}^{a}\sum_{j=1}^{g}\sum_{k=1}^{r}Y_{ijk.}^{2} = Y'Y - \left(\frac{g}{t}\right)Y'_{ijk.}Y_{ijk.}$$

Obsérvese que la suma de la SC UP y SC SUB debe ser igual a la SC total, tanto para la metodología de mínimos cuadrados como para formas cuadráticas o matriciales. Ambas podrían utilizarse para verificar diversos cálculos manuales o para definir una rutina apropiada cuando se apliquen varios paquetes estadísticos.

Uso de InfoGen o InfoStat

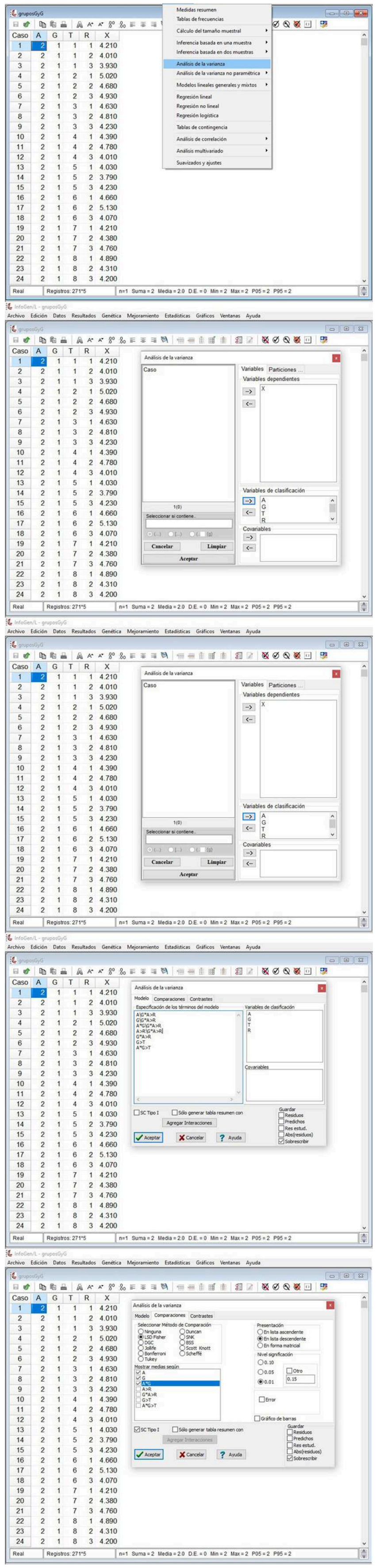
Los rótulos para las columnas serán ambientes, grupos, repeticiones, tratamientos, y variable respuesta, los cuales podrían ser identificados con A, G, R, T, X, respectivamente. Los 270 datos serán capturados en tres grupos, cada uno con 15 variedades, para cada una de sus tres repeticiones, en el mismo orden que Gomez y Gomez (1984) mostró sus datos.

Para construir la base de datos que se muestra en InfoGen, se capturaron algunos datos ficticios con el propósito de 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 deberá generarse en dos etapas: en la primera será obtenido un análisis de varianza general con la partición de efectos en A, G, R(A), AxG, error a, T(G), AxT(G) y error b o residual del modelo.

En la segunda etapa se indicará como realizar un análisis de varianza por grupos de tratamientos o para cada ambiente. Esta misma estrategia será aplicada para obtener las salidas correspondientes a la comparación de medias para cada una de las componentes del modelo lineal. Así, se observará lo siguiente. (1 y 2)

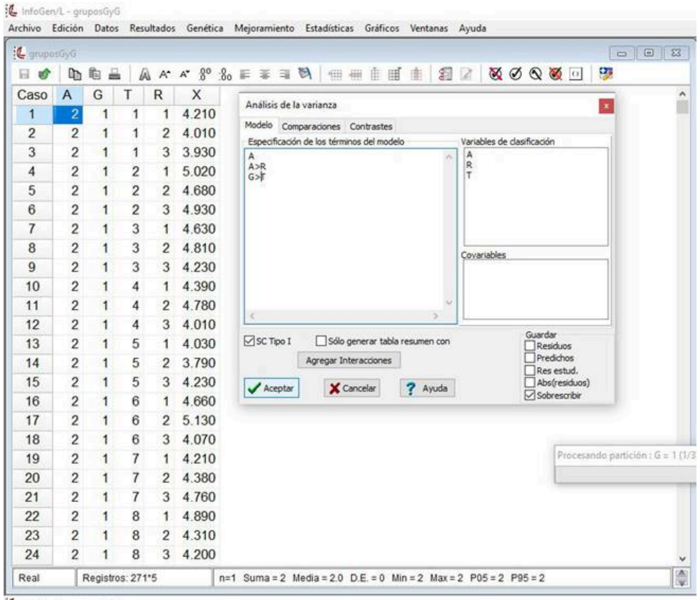
DOI: https://doi.org/10.29312/remexca.v15i7.3831

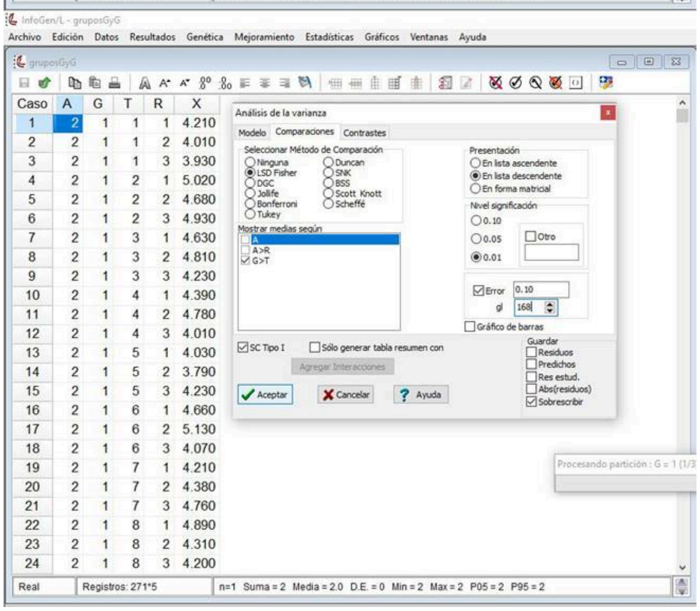
1. Anava general y comparación de medias para los componentes de la unidad principal (UP)



Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i7.3831

Anava y comparación de medias para tratamientos anidados dentro de grupos





DOI: https://doi.org/10.29312/remexca.v15i7.3831

Discusión

Las series de experimentos a través de años, localidades o en ambos, en alguno de los diseños experimentales básicos, han sido de gran relevancia en la investigación agronómica. La construcción de sus modelos lineales para un diseño experimental en bloques completos aleatorizados, la definición de las esperanzas matemáticas de los cuadrados medios y las pruebas de hipótesis apropiadas en el análisis de varianza han sido analizadas y discutidas en Sahagún (1993); Sahagún (1994); Sahagún (2007), entre otros investigadores.

Sahagún (1993) discutió las implicaciones generadas de la aplicación de cuatro modelos lineales para la evaluación de diversos genotipos (G) a través de varios años (A) y localidades (L), o cuando éstos son analizados a través de ambientes, generados con la combinación de los niveles de ambos factores de clasificación, bajo un diseño de bloques completos al azar (DBCA) recomendable para uso en cultivos anuales.

En estos modelos él consideró que los años y las localidades son factores aleatorios y, adicionalmente, definió: en el modelo 1, A, L y G están cruzados y las repeticiones (R) están anidadas dentro de L; en el modelo 2, R está anidado dentro de A y L; en el modelo 3, R está anidado en A y éste último también lo está en L; en el modelo 4, él introdujo el concepto de confusión para los factores A y L, cuya combinación de niveles genera otro factor de clasificación.

En González *et al*. (2008); González *et al*. (2010); González *et al*. (2011), entre otros, se han discutido algunas de las implicaciones agronómicas del uso de una serie de experimentos en DBCA para identificar material genético de alto rendimiento y estabilidad fenotípica cuando se han evaluado un conjunto de variedades e híbridos cuyos progenitores putativos son las razas de maíz Cónico, Chalqueño, Cacahuacintle, Palomero Toluqueño o complejos raciales formados entre algunas de éstas con germoplasma de origen tropical o subtropical, proveniente de líneas endogámicas formadas por el Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) y el Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP), recomendables para siembra comercial en la región central de México.

En este contexto, también ha sido relevante el arreglo de bloques completos balanceados para un diseño de bloques completos al azar, que ha sido analizado y discutido para un solo ensayo por Gomez y Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021); González *et al*. (2024), entre otros, quienes sugirieron que esta agrupación podría realizarse considerando diferencias en altura de planta, ciclo biológico, rendimiento de grano u otra característica cuantitativa importante.

Los cultivares también podrían clasificarse en subconjuntos considerando su origen geográfico y genético, como lo sugirieron González *et al*. (2008); González *et al*. (2011), con y sin arreglo en bloques completos balanceados. En áreas experimentales heterogéneas, como las que predominan en toda la república mexicana, con la propuesta hecha por Gomez y Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021) se probarían las hipótesis estadísticas para los subconjuntos de tratamientos de manera más precisa en comparación con la realizada por González *et al*. (2008, 2010, 2011) u otros investigadores que utilizaron un DBCA sin ABCB.

Debido a la existencia de los errores a y b: la primera sería utilizada para probar las hipótesis relacionadas con los efectos y varianzas para ambientes, grupos, ambientes x grupos y repeticiones dentro de ambientes, mientras que el error b sería utilizada para detectar diferencias significativas entre tratamientos anidados dentro de grupos, así como para la interacción ambientes x tratamientos dentro de grupos.

En el contexto anterior, el error a representa la interacción grupos x repeticiones dentro de ambientes 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 la misma forma como lo plantearon González *et al*. (2024).

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

Revista Mexicana de Ciencias Agrícolas
DOI: https://doi.org/10.29312/remexca.v15i7.3831

de contrastes mutuamente ortogonales, éstos se prueban con el residual del modelo, que es equivalente a su error experimental; la significancia estadística en la prueba de F para tratamientos dentro de grupos depende de si hay o no diferencias estadísticas entre y dentro de grupos en un ABCB-DBCA, tanto para un ensayo como para la serie de experimentos a través de ambientes.

Las notaciones sumatoria y punto han sido de gran utilidad para realizar cálculos manuales en diversas ramas de la estadística y la probabilidad y en el análisis de experimentos agronómicos: su aplicación informal, pero fácil y precisa, se puede consultar en Gomez y Gomez (1984); Martínez (1988); Cochran y Cox (2004), mientras que Mendenhall (1987); Zamudio y Alvarado (1996); Sahagún (1998); Montgonery (2009); Restrepo (2007a, b), entre otros, proponen un uso más formal para evitar confusión en su manejo, particularmente, cuando serán aplicadas las guías para la construcción de modelos fijos, aleatorios, o mixtos, o cuando serán definidas las esperanza matemática de los cuadrados medios para estimar componentes de varianza.

Ambas notaciones también son de gran utilidad para homologar las fórmulas generadas con la técnica de mínimos cuadrados con las derivables a partir de expresiones matriciales o cuadráticas (González *et al*., 2023; González *et al*., 2024). 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; aun cuando Gomez y Gomez (1984); Shikari *et al*. (2015); Maranna *et al*. (2021) no presentaron el modelo lineal correspondiente a un ensayo de un ABCB-DBCA, sus resultados permitieron destacar la relevancia que tiene este tipo de arreglo de unidades experimentales en el diseño y análisis de experimentos agronómicos.

En este contexto, González *et al*. (2024) presentaron información complementaria a la que se encuentra disponible en las publicaciones referenciadas previamente y en el presente estudio, realiza una propuesta para analizar la serie de experimentos a través de ambientes. Los cálculos manuales frecuentemente se consideran como un prerrequisito para la aplicación de software.

En el contexto anterior, en este 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 cuando se incorporen las recomendaciones y sugerencias realizadas por Sahagún (1993); Sahagún (2007), entre otros investigadores.

Si el área experimental en la serie de experimentos conducido en ABCB-DBCA, es dividida en unidad principal (UP) y subunidad (SU), como lo propusieron González *et al*. (2024), sus grados de libertad se calcularían como ar g-1 y ar(t - g), respectivamente, cuya suma origina los art-1 grados de libertad que corresponden al caso anterior y también a una serie de experimentos conducida en un DBCA sin ABCB. Adicionalmente, el total para grados de libertad de tratamientos dentro de grupos sería el mismo que corresponde a cada ensayo individual; es decir:

$$\sum_{i=1}^{g}\left(\frac{t}{g}-1\right) = t-g.$$

Así, será más fácil calcular los grados de libertad para el error b. González *et al*. (2019) fraccionaron los efectos entre tratamientos en grupos en un DBCA sin ABCB aplicando la técnica de contrastes mutuamente ortogonales pero la precisión con la que se prueban las hipótesis estadísticas de interés, en un área experimental más heterogénea, podrían más confiables utilizando un ABCB-DBCA.

Para verificar los cálculos relacionados con las sumas de cuadrados (SC) que serán generados en el análisis de varianza en una serie de experimentos con y sin ABCB-DBCA, deberán compararse las salidas que genera este diseño experimental en ambos tipos de arreglos de unidades experimentales. La SC T(G) más la SC G debe ser igual a la SC de T. También, la SC A*T(G) más SC A*G será igual a SC T*A y finalmente, la SC del error experimental será igual a la adición de las SC de los errores a y b.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

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, b); Piepho *et al*. (2003). 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, las componentes de la UP deberán probarse utilizando el error a, mientras que las correspondientes a la SUB, emplearán el error b. 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 y adicionalmente, ambos permiten realizar la corrección que debe efectuarse a la diferencia mínima significativa honesta o prueba de Tukey, pero deben capturarse manualmente los grados de libertad y el cuadrado medio del error b, generados con todos los datos registrados en la serie de experimentos; los correspondientes a cada ensayo también serán utilizados para realizar este tipo de pruebas de manera independiente.

Si en la serie de experimentos las diferencias entre grupos de tratamientos no son significativas, InfoGen o InfoStat pueden generar un análisis de varianza y una comparación de medias con la 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 considerado en este estudio se construyó considerando que los ambientes y los grupos están cruzados, y que las repeticiones y los tratamientos están anidados dentro de ambientes y dentro de grupos, respectivamente. Las fórmulas para calcular grados de libertad y sumas de cuadrados se simplificarán al dividir el área experimental en unidad principal y subunidad: en ambos se alojan los errores a y b, respectivamente; la primera es la interacción grupos x repeticiones dentro de ambientes, y la segunda es el residual del modelo.

La técnica de mínimos cuadrados es más fácil de aplicar cuando se usa un paquete estadístico, especialmente si el número de experimentos y de variables por analizar es grande. InfoGen y InfoStat, son muy flexibles cuando se aplica la prueba de Tukey a los tratamientos anidados dentro de grupos, promediando los valores sobre ambientes y repeticiones, 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 en el ABCB-DBCA son iguales estadísticamente, los datos podrían analizarse como una serie de experimentos en DBCA utilizando el mismo archivo; como opción para generar los mismos resultados puede recurrirse al software OPSTAT, disponible gratuitamente en su sitio WEB, o a cualquier otro paquete estadístico, como SAS o Agrobase, entre otros.

DOI: https://doi.org/10.29312/remexca.v15i7.3831

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. Editorial Brujas. Córdoba, Argentina. 348 p.

3 Cochran, W. G. y Cox, G. M. 2004. Diseños experimentales. Editorial Trillas, SA. de CV. 6ta. Reimpresión. México, DF. 661 p.

4 Di Rienzo, J. A.; Casanoves, F.; Balzarini, M. G.; González, L.; Tablada, M. and Robledo, C. W. 2008. InfoStat Versión 2008. Grupo InfoStat, FCA. Universidad Nacional de Córdoba. Argentina. (https://www.infostat.com.ar).

5 Gomez, K. A. and 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, Atlacomulco, México. Revista Fitotecnia Mexicana. 31(1):67-76.

7 González, H. A.; Pérez, D. J.; Sahagún, C. J.; Franco, O.; Morales, E. J.; Rubí, A. M.; Gutiérrez, F. y Balbuena, A. 2010. Aplicación y comparación de métodos univariados para evaluar la estabilidad en maíces del Valle de Toluca, Atlacomulco, 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. 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 González, H. A.; Pérez, L. D. J.; Hernández, A. J.; Franco, M. J. R.; Rubí, A. M. y Balbuena, M. A. 2024. Tratamientos anidados dentro de grupos en arreglo de bloques completos balanceados. Revista Mexicana de Ciencias Agrícolas. 15(2)e3634.

12 Little, T. M. y Hills, F. J. 2008. Métodos estadísticos para la investigación en la agricultura. Editorial Trillas SA. de CV. México. 270 p.

13 Maranna, S.; Nataraj, V.; Kumawat, G.; Chandra, S.; Rajesh, V.; Ramteko, R.; Manohar, P. R.; Ratnaparkhe, M. B.; Husain, S. M.; Gupta, S. and Khandekar, N. 2021. Breeding for higher yield, early maturity, wider adaptability and waterlogging tolerance in soybean (Glycine max L.): a case study. Scientific reports. 11:22853. https://doi.org/10.1038/s41598-021-02064-x.

14 Martínez, G. A. 1988. Diseños experimentales. Métodos y elementos de teoría . Editorial Trillas, 1ra. Ed. México. 756 p.

15 Mendenhall, W. 1987. Introducción a la probabilidad y la estadística. Grupo Editorial Iberoamérica. 1ra. Ed. México. 626 p.

16 Montgomery, D. C. 2009. Design and analysis of experiments. 7th Ed. John Wiley (Sons, Inc. USA. 656 p.

Revista Mexicana de Ciencias Agrícolas
REMEXCA
DOI: https://doi.org/10.29312/remexca.v15i7.3831

17 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.

18 Restrepo, L. F. 2007a. Diagramas de estructuras en el análisis de varianza. Revista Colombiana de Ciencias Pecuarias. 20(1):202-208.

19 Restrepo, B. L. F. 2007b. La esperanza del cuadrado medio. Revista Colombiana de Ciencias Pecuarias. 20(2):193-201.

20 Sahagún, C. J. 1993. Funcionalidad de cuatro modelos para las evaluaciones genotípicas en series de experimentos. Revista Fitotecnia Mexicana. 16(3):161-171.

21 Sahagún, C. J. 1994. Evaluación de genotipos en series de experimentos: diferencias en parámetros genéticos generados en dos modelos. Revista Fitotecnia Mexicana. 17(2):116-125.

22 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. Boletín técnico núm. 2. 64 p.

23 Sahagún, C. J. 2007. Evaluación de genotipos en heterogeneidad meteorológica intrarregional: confusión *vs* anidamiento de años en localidades. Revista Fitotecnia Mexicana. 30(1):97-104.

24 Sahagún, C. J. 2007. Estadística descriptiva y probabilidad: una perspectiva biológica. 2[da]. Ed. Universidad Autónoma Chapingo, México. 282 p.

25 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 DS. Hooda (RC. Hasija Department of Mathematics Statistics, CCS HAU, Hisar. 139-143 pp.

26 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.

27 Zamudio, S. F. J. y Alvarado, S. A. A. 1996. Análisis de diseños experimentales con igual número de submuestras. 1[ra]. Ed. División de Ciencias Forestales. Universidad Autónoma Chapingo. Texcoco, Estado de México, México. 85 p.

Serie de experimentos para tratamientos anidados en grupos con arreglo 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 April 2024
Date accepted: 01 August 2024
Publication date: 09 December 2024
Publication date: Oct-Nov 2024
Volume: 15
Issue: 7
Electronic Location Identifier: e3831
DOI: 10.29312/remexca.v15i7.3831

Categories

Subject: Artículo

Palabras clave:

Palabras clave:

diseño de bloques completos al azar

InfoStat

suma de cuadrados y formulas cuadráticas.

Counts

Figures: 2

Tables: 0

Equations: 18

References: 27

Pages: 0