

Submuestreo balanceado en experimentos monofactoriales usando InfoStat e InfoGen: validación con SAS

Andrés González-Huerta^{1§}
Delfina de Jesús Pérez-López¹
Artemio Balbuena-Melgarejo¹
J. Ramón Pascual Franco-Martínez¹
Francisco Gutiérrez-Rodríguez¹
José Antonio Rodríguez-González²

¹Centro de Investigación y Estudios Avanzados en Fitomejoramiento-Facultad de Ciencias Agrícolas-Universidad Autónoma del Estado de México. *Campus* Universitario 'El Cerrillo', El Cerrillo Piedras Blancas, Toluca, Estado de México, México. AP. 435. Tel. 722 2965531, ext. 148. (djperezl@uaemex.mx; abalbuenam@uaemex.mx; jrfrancom@uaemex.mx; fgfca@yahoo.com) ²Posgrado en Ciencias Agropecuarias y Recursos Naturales-Universidad Autónoma del Estado de México. (rodyof07@outlook.com).

§Autor para correspondencia: agonzalezh@uaemex.mx.

Resumen

Aun en la actualidad existe poca información publicada con relación al análisis de experimentos monofactoriales cuando se usa igual número de submuestras dentro de cada unidad experimental. En este estudio se analizan datos de floración masculina registrados en cuatro variedades de maíz (*Zea mays* L.) establecidas en condiciones de campo usando cuatro repeticiones por tratamiento, dentro de cada unidad experimental se registraron 30 datos, para el presente estudio sólo son considerados tres de éstos. Los diseños experimentales elegidos fueron completamente al azar, bloques completos al azar y cuadro latino. Las salidas fueron obtenidas con InfoStat y corresponden a un análisis de varianza y a una comparación de medias de tratamientos con la prueba de Tukey ($p=0.01$), estas también se pueden generar con InfoGen aplicando el mismo procedimiento. Los datos que conducen a ambos resultados fueron usados para cálculos manuales y los resultados son validados con el sistema para análisis estadístico. Debido que los datos son los mismos el error muestral es común en los tres diseños experimentales y se muestra cómo obtener el error conjunto, la diferencia entre ambos genera el error experimental. Para simplificar el procedimiento en la computadora personal se elabora una sola base de datos. Sólo para el diseño en cuadro latino se proporcionan las expresiones matriciales que permiten homologar el cálculo manual con sumas de cuadrados en el análisis de varianza. Si el objetivo secundario fuera comparar los tres diseños experimentales el soporte estadístico generado por éstos lo permitiría, en una sola corrida usando SAS e individualmente para cada diseño aplicando InfoStat y InfoGen.

Palabras clave: algebra matricial, error muestral en diseños experimentales, paquetes estadísticos gratuitos, suma de cuadrados en ensayos con submuestreo.

Recibido: diciembre de 2022

Aceptado: febrero de 2023

Introducción

Cuando se aplica un diseño experimental en las ciencias agropecuarias y forestales se usan totales o promedios aritméticos para probar hipótesis estadísticas relacionadas con las preguntas que surgen con relación a la estructura de tratamientos que está siendo evaluada (Zamudio y Alvarado, 1996; Sahagún, 1998; Restrepo, 2007a, 2007b). A través de los análisis de varianza se hace una partición de efectos o varianzas relacionadas con las fuentes de variabilidad que están implícitas en los modelos genético-estadísticos que son de interés para los usuarios (Sahagún, 1991; Sahagún, 1998; Piepho *et al.*, 2003; Restrepo, 2007a, 2007b).

En especies anuales, como cereales de grano pequeño, frecuentemente se miden y cuantifican más de dos plantas o varias partes de éstas dentro de cada unidad experimental para mejorar o incrementar la precisión con la que se prueban dichas hipótesis, el tamaño de la muestra, más repeticiones, tratamientos o ambos, así como un mejor control local en el área experimental, entre otros, contribuyen a dicho propósito.

En esta situación los usuarios podrán usar cada una de las observaciones disponibles dentro de cada parcela o unidad experimental (Gomez y Gomez, 1984; Martínez, 1988; Zamudio y Alvarado, 1996). En libros, artículos, tesis, folletos técnicos o en otras fuentes de información confiable se ha observado que existe poca información cuando se realiza submuestreo dentro de las unidades experimentales, especialmente cuando se aplica algún paquete estadístico (Martínez, 1988; Freund y Wilson, 1993; Zamudio y Alvarado, 1996). Como las sumas de cuadrados se pueden obtener con dos metodologías, el otro inconveniente sería presentar la homologación entre las fórmulas que las generan, especialmente cuando éstas son más complejas.

Zamudio y Alvarado (1996) presentaron un algoritmo para realizar submuestreo en diseños experimentales mediante la aplicación de álgebra matricial, ellos enfatizaron el hecho de elegir el modelo estadístico correcto y dar prioridad a la variabilidad dentro de la unidad experimental, la variabilidad que existe en cada una de éstas últimas tiene dos componentes: una fija y otra aleatoria. El análisis estadístico se realizó en dos etapas a través de códigos para SAS, para analizar independientemente, los datos para generar un análisis de varianza en los diseños experimentales completamente al azar (DCA), bloques completos al azar (DBCA) y cuadro latino (DCL).

En el presente estudio el objetivo principal fue aplicar InfoStat y InfoGen para analizar datos provenientes de un experimento con submuestreo balanceado dentro de cada unidad experimental en los diseños DCA, DBCA y DCL, el objetivo secundario fue validar los cálculos manuales y las salidas generadas con ambos paquetes estadísticos usando SAS.

Materiales y métodos

Material biológico

En este estudio fueron consideradas cuatro variedades de maíz (*Zea mays* L.) evaluadas en campo en el año 2010 en un terreno de la Facultad de Ciencias Agrícolas de la UAEMéx: el cv. Ixtlahuaca, raza Cónico, una variedad de la raza Cacahuacintle, una población nativa de la raza Palomero Toluqueño y el híbrido Cóndor. Estos y otros materiales de maíz han sido evaluados por González *et al.* (2008, 2010), pero los datos considerados en esta investigación no fueron publicados.

Diseño experimental y tamaño de la parcela

El ensayo fue sembrado en campo en un diseño experimental Cuadro Latino 4 x 4. Cada parcela experimental (UE) constó de 6 m de longitud, con separación entre surcos de 0.8 m. Hubo tres hileras de plantas por cada UE y dentro de cada una de éstas últimas se registraron datos en 30 plantas, pero sólo serán consideradas tres observaciones.

Modelos estadísticos y diseños experimentales

Para la construcción de los modelos convencionales, así como los que incluyen submuestreo, que se describen a continuación, pueden consultarse las guías proporcionadas por Sahagún (1998); Piepho *et al.* (2003); Restrepo (2007a, 2007b). Los modelos son:

Para DCA: $Y_{ikl} = \mu + \tau_k + \delta_{kl} + \varepsilon_{ikl}$

Para DBCA: $Y_{ikl} = \mu + H_i + \tau_k + \delta_{kl} + \varepsilon_{ikl}$

Para DCL: $Y_{ijkl} = \mu + H_i + C_j + \tau_k + \delta_{kl} + \varepsilon_{ijkl}$

Donde: Y= la variable respuesta; μ = la media general; τ_k = efecto que causa la k-ésima variedad; H_i , C_j = la heterogeneidad ambiental que existe entre hileras y entre columnas. Los δ 's y los ε 's son los errores muestral y experimental, respectivamente: ambos determinarán el error conjunto. En Zamudio y Alvarado (1996) se describen los dos primeros modelos y cómo estimar sus componentes con álgebra matricial.

Procedimientos aritmético-algebráico-matriciales

Análisis de varianza (Anava)

Las etapas (E's) que permitirán la verificación de los cálculos manuales son: E1) concentrar los datos como se muestra en el Cuadro 1, obtener subtotales y totales.

Cuadro 1. Datos para floración masculina considerando el número de observaciones registradas en cada tratamiento (número entre paréntesis), en cada combinación de hilera y columna.

Hilera (i)	Muestra (l)	Columnas (j)				Total
		1	2	3	4	
1	1	95	85	93	105	
1	2	96(3)	86(4)	95(2)	104(1)	
1	3	98	83	96	105	
		289	254	284	314	1 141
2	1	85	92	104	94	
2	2	83(4)	94(3)	105(1)	92(2)	
2	3	84	96	106	95	
		252	282	315	281	1 130
3	1	93	104	87	95	

Hilera (i)	Muestra (l)	Columnas (j)				Total
		1	2	3	4	
3	2	90(2)	107(1)	85(4)	94(3)	
3	3	92	106	86	96	
		275	317	258	285	1 135
4	1	106	92	95	87	
4	2	105(1)	93(2)	93(3)	84(4)	
4	3	104	94	97	87	
		315	279	285	258	1 137
Subtotal		1 131	1 132	1 142	1 138	4 543

Totales para: 1= Córdor= 1261; 2= Ixtlahuaca=1119; 3= Cacahuacintle = 1141; 4= Palomero Toluqueño=1022.

E2) elaborar el formato del Anava con submuestreo; E3) calcular grados de libertad (GL). En un DCL, el número de tratamientos (T), repeticiones (R), hileras (H) y columnas (C) es igual; ningún T debe repetirse en H o C. Si S es el tamaño de muestra registrado dentro de cada unidad experimental, entonces: T= R= H= C= 4 y S= 3. Los GL comunes en los tres diseños experimentales se calculan como: GL total= trs- 1= ths - 1 = tcs - 1= 4(4)(3) - 1= 47; GL T= GL R = GL H= GL C= t - 1= r -1= h - 1= c - 1= 4-1= 3; GL error muestral (EM)= tr (s-1)= th(s-1)= tc (s-1) = 4(4)(3-1) = 32.

Ahora se calculará lo que depende del diseño experimental elegido. El error conjunto (EC) es la suma del EM y del error experimental, éste último será identificado como EE, se obtendrá lo siguiente:

DCA: GL EC= GL Total - GL T= 47 - 3= 44

GL EE= t(r-1)= t(h-1)= t(c -1)= 4(4-1)= 12

Para verificación: GL EC= GL EM + GL EE, así: GL EC= 32 + 12= 44

DBCA: GL EC= GL total - GL R - GL T= 47 - 3 - 3= 41

GL EE= (t-1)(r-1)= (t-1)(h-1)= (t-1)(c-1)= 3(3)= 9. Para verificación: GL EC= GL EM + GL EE= 32 + 9= 41

DCL: GL EC= GL Total - GL H - GL C - GL T = 47 - 3 - 3 - 3= 38

GL EE= (t-1)(t-2)= (h-1)(h-2)= (c-1)(c-2)= (4-1)(4-2)= 6. Para verificación: GL EC= GL EM + GL EE= 32 + 6= 38

E4) calcular suma de cuadrados (SC). Primero se calculan las SC que pueden homologarse en los tres diseños experimentales. Factor de corrección (FC)= $\frac{(\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^s Y_{ijkl})^2}{hts} = \frac{(95+95+95+...+87)^2}{4(4)(3)} = 429\,976.02$; SC Total= $\sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^s Y_{ijkl}^2 - FC = (95^2+95^2+95^2+...+87^2) - \frac{(95+95+95+...+87)^2}{4(4)(3)} = 432\,483 - 429\,976.02 = 2506.98$.

Al igual que en la sección anterior, ahora H o C podría considerarse como R en un DBCA, en el denominador de las fórmulas que se muestran en esta sección uno de éstos será nulo, pero en las sumatorias éstas serán consideradas según procedan. Por esta razón, en el denominador del FC se incluye h o c, pero no ambas.

$$SC T = \frac{\sum_{k=1}^t Y_{..k}^2}{hs} - FC = \frac{(1261^2 + 1119^2 + 1141^2 + 1022^2)}{4(3)} - \frac{(95+95+95+\dots,+87)^2}{4(4)(3)} = 432\,387.25 - 429\,976.02 = 2411.23;$$

$$SC R = SC H = \frac{\sum_{i=1}^h Y_{i..}^2}{ts} - FC = \frac{(1141^2 + 1130^2 + 1135^2 + 1137^2)}{4(3)} - \frac{(95+95+95+\dots,+87)^2}{4(4)(3)} = 429\,981.249 - 429\,976.02 = 5.2291;$$

$$SC EM = \sum_{i=1}^h \sum_{j=1}^c \sum_{k=1}^t \sum_{l=1}^s Y_{ijkl}^2 - \frac{\sum_{i=1}^h \sum_{k=1}^t Y_{i.k}^2}{s}$$

Para calcular esta SC debe generarse el (Cuadro 2).

Cuadro 2. Subtotales y totales para calcular SC EM.

Hilera (i)	Tratamiento (k)				Total
	1	2	3	4	
1	314	284	289	254	1 141
2	315	281	282	252	1 130
3	317	275	285	258	1 135
4	315	279	285	258	1 137
Total	1 261	1 119	1 141	1 022	4 543

Con los datos contenidos en los Cuadros 1 y 2 se estima: $SC EM = (95^2 + 95^2 + 95^2 + \dots + 87^2) - \frac{(314^2 + 284^2 + 289^2 + \dots + 258^2)}{3} = 432\,483 - 432\,420.333 = 62.666$. Ahora se calcula lo que es específico para cada diseño experimental. Para DCL: $SC H = SC R$; éste se calculó previamente. Ya se comentó que $SC R$ también podría considerarse como $SC C$.

$$SCC = \frac{\sum_{j=1}^c Y_{.j.}^2}{ts} - FC = \frac{(1131^2 + 1132^2 + 1142^2 + 1138^2)}{4(3)} - \frac{(95+95+95+\dots,+87)^2}{4(4)(3)} = 429\,982.75 - 429\,976.02 = 6.73;$$

$$EC = SC Total - SC H - SC C - SC T = 2506.98 - 5.2291 - 6.73 - 2411.23 = 83.7909$$

Como: $SC EC = SC EE + SC EM$, entonces:

$$SC EE = SC EC - SC EM = 83.7909 - 62.666 = 21.1249$$

$$\text{Para DBCA: } SC EC = SC Total - SC H - SC T = 2506.98 - 5.2291 - 2411.23 = 90.5209$$

$$\text{Por lo tanto: } SC EE = SC EC - SC EM = 90.5209 - 62.67 = 27.8509$$

$$\text{Para: DCA: } SC EC = SC Total - SC T = 2506.98 - 2411.23 = 95.75$$

$$\text{Así: } SC EE = SC EC - SC EM = 95.75 - 62.67 = 33.08.$$

E5. Los cálculos restantes y las pruebas de hipótesis para H, C y T se obtienen de manera convencional al caso cuando se consideran promedios aritméticos por unidad experimental (ver las salidas de InfoStat).

Método matricial

$$SC Total = Y'Y - \left(\frac{1}{hts}\right) Y'JY$$

$$SC H = \left(\frac{1}{ts}\right) (Y'_{i..})(Y_{i..}) - \left(\frac{1}{hts}\right) Y'JY$$

$$SC C = \left(\frac{1}{ts}\right) (Y'_{.j.})(Y_{.j.}) - \left(\frac{1}{hts}\right) Y'JY$$

$$\begin{aligned}
 SC\ T &= \left(\frac{1}{hs}\right) (Y' \dots_k)(Y \dots_k) - \left(\frac{1}{hts}\right) Y' J Y \\
 SC\ EM &= Y' Y - \left(\frac{1}{s}\right) (Y'_{i \cdot k})(Y_{i \cdot k}) \\
 SC\ EC &= Y' Y - \left(\frac{1}{ts}\right) (Y'_{i \dots})(Y_{i \dots}) - \left(\frac{1}{ts}\right) (Y'_{\cdot j \dots})(Y_{\cdot j \dots}) - \left(\frac{1}{hs}\right) (Y' \dots_k)(Y \dots_k) + \left(\frac{2}{hts}\right) Y' J Y \\
 SC\ EE &= \left(\frac{1}{s}\right) (Y'_{i \cdot k})(Y_{i \cdot k}) + \left(\frac{2}{hts}\right) Y' J Y - \left(\frac{1}{ts}\right) (Y'_{i \dots})(Y_{i \dots}) - \left(\frac{1}{ts}\right) (Y'_{\cdot j \dots})(Y_{\cdot j \dots}) - \left(\frac{1}{hs}\right) (Y' \dots_k)(Y \dots_k).
 \end{aligned}$$

Para simplificar cálculos, SC EE= SC EC - SC EM. En las matrices Y, Y' se capturan las tres observaciones del tratamiento 1 en la repetición 1, las tres observaciones de éste en su repetición 2 y así sucesivamente. Es decir:

$$Y_{1 \times 48} = \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix}$$

Y' es la transpuesta de la matriz Y; la matriz J, de 1's, tiene 48 hileras y 48 columnas. Las matrices restantes se construyen con totales; su o sus subíndices indican los totales que deben introducirse dentro de ellas. En esta sección podría ser de utilidad consultar las publicaciones de Jasso *et al.* (2022); Pérez *et al.* (2022), en las cuales se realizan cálculos matriciales, como los que a continuación serán mostrados, con un enfoque estadístico - genético. Así: SC total = Y'Y - $\left(\frac{1}{hts}\right)$ Y'JY.

$$\begin{aligned}
 &= [105\ 104\ 105\ \dots\ 87] \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix} - \left(\frac{1}{4(4)(3)}\right) [105\ 104\ 105\ \dots\ 87] \begin{bmatrix} 1\ 1\ 1 & \dots & 1\ 1\ 1 \\ \vdots & \ddots & \vdots \\ 1\ 1\ 1 & \dots & 1\ 1\ 1 \end{bmatrix} \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix} = \\
 &= 432\ 483 - 429\ 976.02 = 2506.98
 \end{aligned}$$

$$SC\ Trat = \left(\frac{1}{hs}\right) (Y' \dots_k)(Y \dots_k) - \left(\frac{1}{hts}\right) Y' J Y$$

$$= \left(\frac{1}{4(3)}\right) [1261 \ 1119 \ 1141 \ 1022] \begin{bmatrix} 1261 \\ 1119 \\ 1141 \\ 1022 \end{bmatrix} - \left(\frac{1}{4(4)(3)}\right) [105 \ 104 \ 105 \ \dots \ 87] \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ \vdots & & \ddots & & \vdots & & \\ 1 & 1 & 1 & \dots & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 105 \\ 104 \\ 105 \\ 104 \\ 105 \\ 106 \\ \dots \\ 87 \\ 84 \\ 87 \end{bmatrix}$$

$$= 432 \ 387.25 - 429 \ 976.02 = 2411.23$$

En la primera matriz se capturaron totales, en el orden primero, segundo, tercero y cuarto tratamiento. El mismo orden debe mantenerse para hileras y columnas. La expresión común, restada en los dos cálculos anteriores, corresponde al factor de corrección. En esta sección, como se sugirió en Jasso *et al.* (2022); Pérez *et al.* (2022), también será de gran utilidad el uso de una calculadora de matrices, disponible gratuitamente en su sitio web: (<https://matrixcalc.org>).

Base de datos y análisis estadístico

En InfoStat y InfoGen se elabora una tabla rotulando verticalmente con H, C, T, S, agh, para identificar hileras, columnas, tratamientos, submuestra y variable respuesta, respectivamente. Para cada combinación de hilera, columna y tratamiento se capturan cíclicamente los valores de las muestras 1, 2, 3. Después de verificar la base de datos y realizar su resguardo, en menú principal elegir: ‘estadísticas/análisis de la varianza’. Con flecha adelante, ‘agh’ será enviado a variables dependientes y ‘H, C, T y S’ serán trasladadas a variables de clasificación. elegir ‘aceptar’. ‘H, C, T y H*C*T’ serán capturadas verticalmente en ‘especificación de los términos del modelo’, en las tres primeras, después de la diagonal inversa (\) debe escribirse H*C*T, que es el error experimental. En variables de clasificación debe mostrarse ‘H, C, T y S’.

Resultados

Con el procedimiento anterior se generará el análisis de varianza para un DCL con submuestreo. En las (Figuras 1, 2, 3, 4) que muestran el procedimiento anterior.

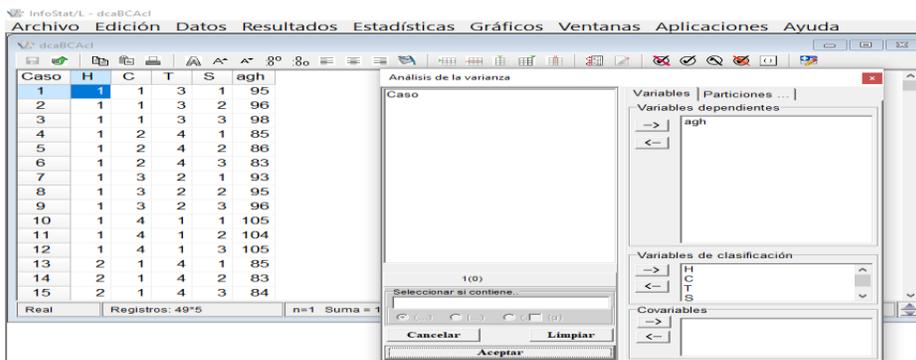


Figura 1.

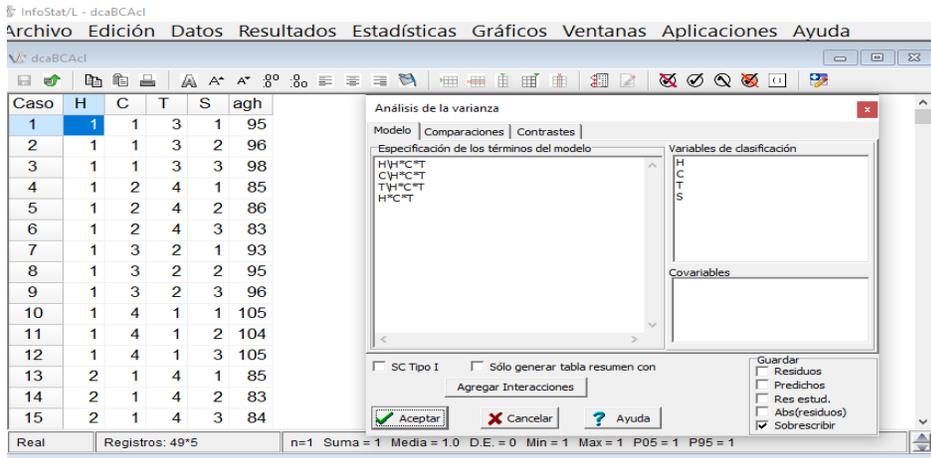


Figura 2.

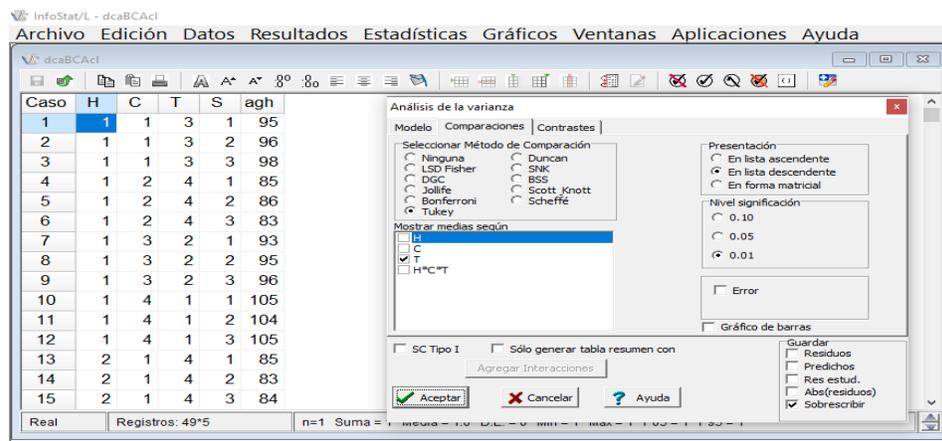


Figura 3.

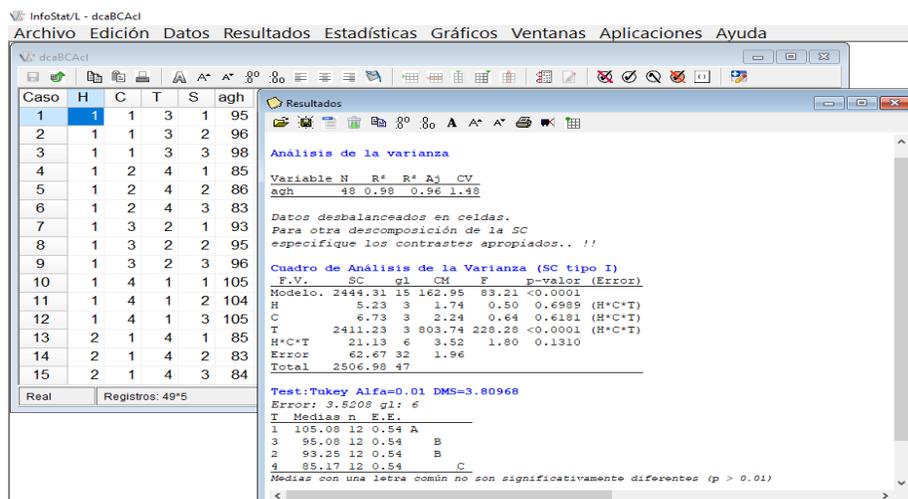


Figura 4.

Para generar el Anava para un DBCA con submuestreo, ‘agh’ se enviará a variables dependientes y en variables de clasificación deberá mostrarse ‘H, T, S’; elegir aceptar. En ‘especificación de los términos del modelo’ deberá capturarse verticalmente ‘H, T y H*T’, en las dos primeras, después del signo \, debe escribirse H*T, que es el error experimental. En variables de clasificación el software mostrará ‘H, T y S’. El procedimiento se muestra en las (Figuras 5, 6, 7, 8).

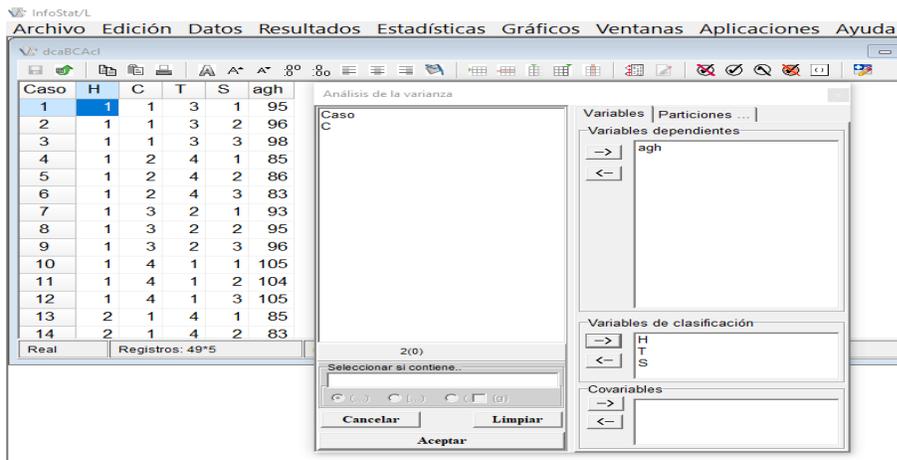


Figura 5.

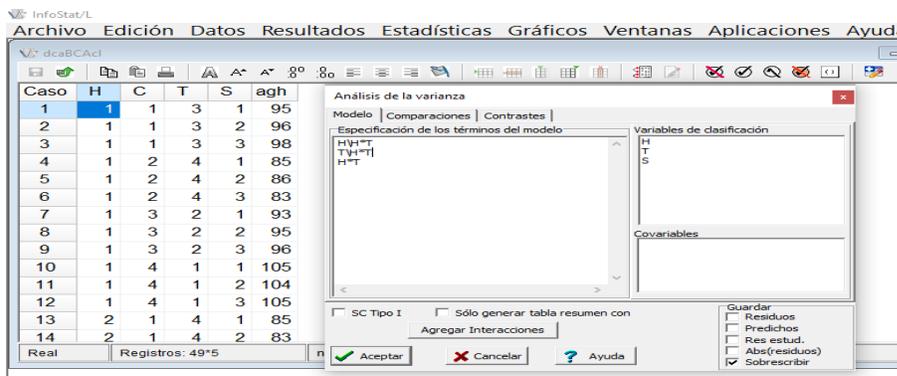


Figura 6.

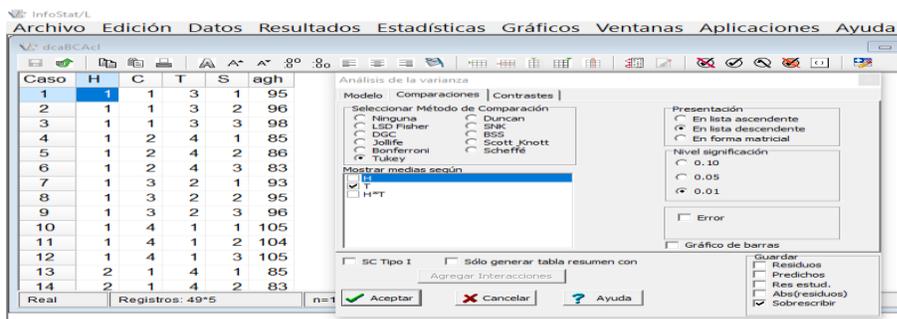


Figura 7.

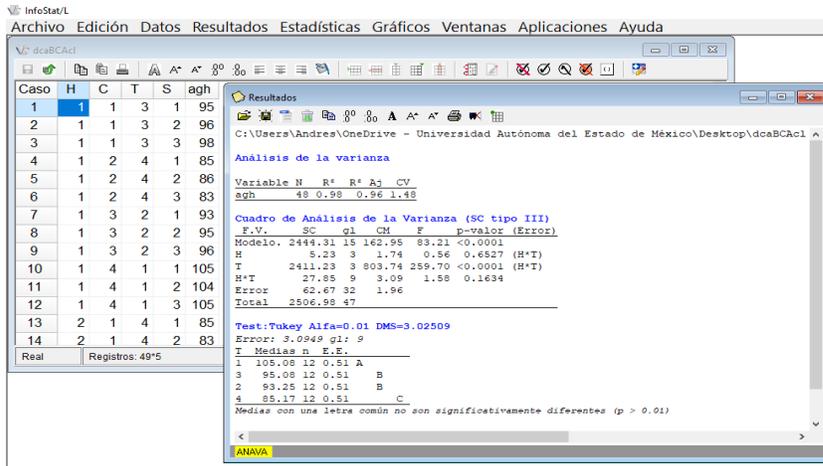


Figura 8.

Para obtener el Anava para un DCA con submuestreo, 'agh' se enviará a variables dependientes y en variables de clasificación deberá mostrarse 'H, C, T'. Al 'aceptar', se mostrará verticalmente, en 'especificación de los términos del modelo', lo siguiente: 'T\H*C*T' y 'H*C*T'. Las variables de clasificación serán 'H, C, T'. Elegir 'aceptar'. En los tres diseños experimentales el software calculará por default el residual del modelo o error muestral (Figuras 9, 10, 11, 12).

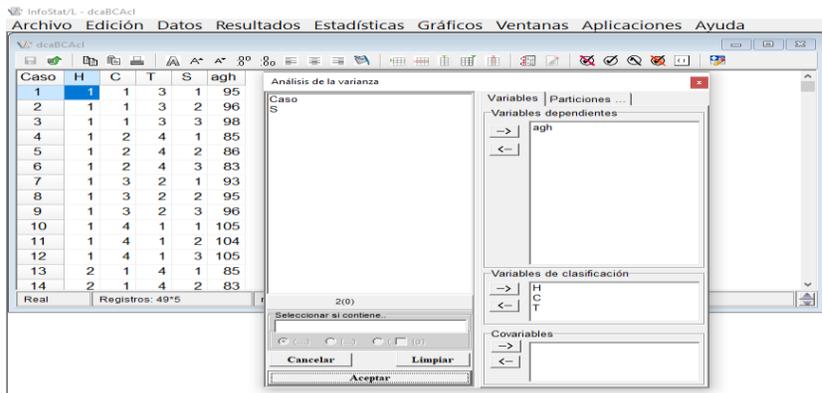


Figura 9.

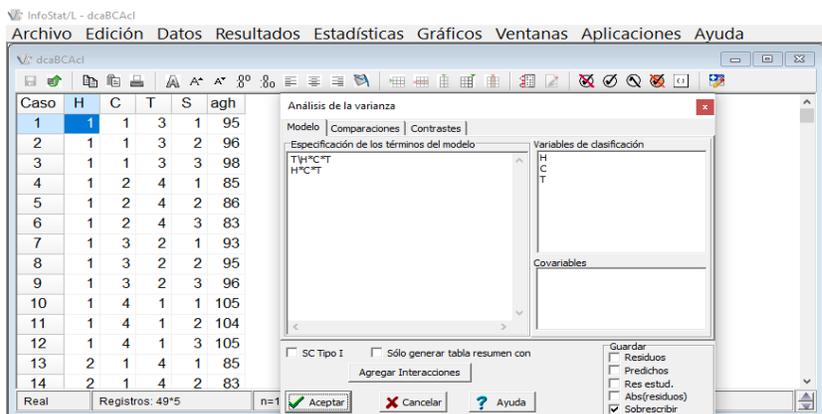


Figura 10.

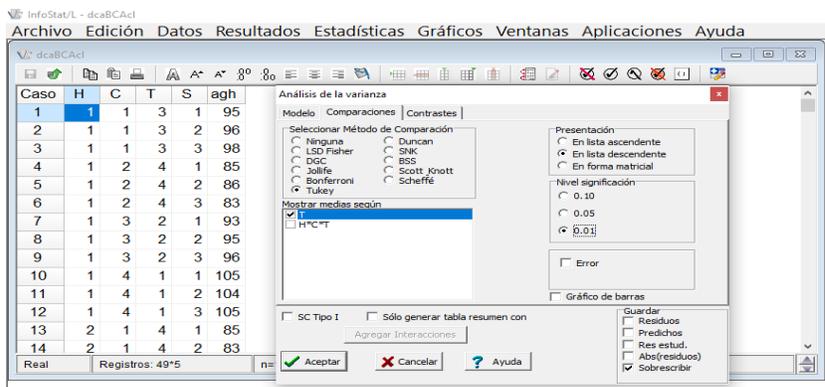


Figura 11.

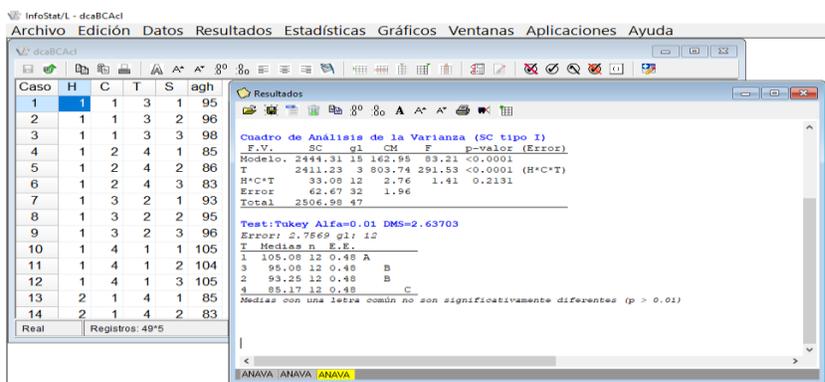


Figura 12.

Validación de resultados con SAS

Desde el programa editor de este software se captura: Title 1 ‘submuestreo dentro de las unidades experimentales en tres diseños’; Title 2 ‘anava y comparación de medias de tratamientos’; data agh22; input H C T S Z; cards;

1 1 3 1 95
1 1 3 2 96
1 1 3 3 98

...

4 4 4 3 87

;

Proc GLM; Class H C T; Model Z= H C T H*C*T/ss3; Test h= H C T e=H*C*T; Means T/Tukey lines alpha=0.01 e= H*C*T; Run;Quit;

Proc GLM; Class H T; Model Z= H T H*T/ss3; Test h= H T e= H*T; Means T/Tukey lines alpha=0.01 e= H*T; Run;Quit;

Proc GLM; Class H C T; Model Z= T H*C*T/ss3; Test h= T e= H*C*T; Means T/Tukey lines alpha= 0.01 e= H*C*T; Run;Quit;

Discusión

El análisis de varianza (Anava) y la comparación de medias de tratamientos son dos metodologías de gran utilidad en el diseño y análisis de experimentos; la primera siempre condiciona el uso de la segunda. El Anava permite probar hipótesis estadísticas con relación a las componentes de naturaleza fija, aleatoria o mixta que conforman los modelos matemáticos que frecuentemente se usan en las ciencias agropecuarias y forestales, entre otras; la variabilidad total que es medida en cada una de las variables de interés es fraccionada en efectos y varianzas para cada una de sus componentes (Martínez, 1988; Sahagún, 1998; Piepho *et al.*, 2003; Restrepo, 2007a, 2007b).

Las comparaciones múltiples de medias de tratamientos, el contraste entre un testigo o control *versus* los $t-1$ tratamientos restantes, la subdivisión de la variabilidad contenida en la estructura de tratamientos en contrastes o polinomios ortogonales, la aplicación de técnicas univariadas o multivariadas, como los análisis de regresión o de componentes principales, sólo es justificable si por medio del Anava se rechazan las hipótesis nulas que fueron planteadas *a priori* o *a posteriori* (Sahagún, 1991; Di Rienzo *et al.*, 2008; Balzarini *et al.*, 2008; Balzarini *et al.*, 2016).

El submuestreo en diseños experimentales también se fundamenta en la aplicación de un Anava (Gomez y Gomez, 1984; Martínez, 1988; Zamudio y Alvarado, 1996; Hansen *et al.*, 2006). Para simplificar los cálculos, el error conjunto que es considerado en el presente estudio, que es el residual en los tres modelos estadísticos descritos, es dividido en error experimental y error muestral, en los diseños experimentales DCA, DBCA y DCL conduce a pruebas de hipótesis asociadas a efectos y varianzas que involucran a ambos tipos de error, con y sin submuestreo.

En el presente estudio, los procedimientos delineados para InfoStat y InfoGen, así como los propios que permiten la validación de los cálculos manuales y las salidas que se generan con ambos softwares o con SAS, son fáciles de implementar en sus plataformas y resultan confiables para analizar individualmente cada uno de los tres diseños experimentales o para analizarlos en una sola corrida, tanto convencionalmente como por medio de submuestreo. En Balzarini *et al.* (2008); Di Rienzo *et al.* (2008); Balzarini *et al.* (2016), están descritos esos procedimientos para obtener un Anava, una comparación de medias de tratamientos, o para aplicar contrastes ortogonales cuando se emplea InfoStat y InfoGen, pero no lo están si se considera submuestreo dentro de la unidad o parcela experimental. En este contexto es que se utiliza SAS para la validación de los resultados que generan ambos paquetes estadísticos.

Gómez y Gómez (1984) mostraron los cálculos para submuestreo en un DBCA, a partir de los cuadrados medios del Anava ellos estimaron las varianzas del error experimental y muestral y calcularon los coeficientes de variación para la variable respuesta usando ambas varianzas. Freund y Wilson (1993) también mostraron una salida generada con SAS para el caso de un DBCA.

Los códigos y procedimientos que son usados en el presente estudio validaron correctamente los resultados que ellos mostraron con relación al Anava, aunque en la salida de SAS que presentaron Freund y Wilson (1993) las pruebas de F para repeticiones, tratamientos y error experimental se hicieron utilizando el cuadrado medio del error muestral, por lo que de acuerdo con Martínez (1988); Zamudio y Alvarado (1996); Sahagún (1998), éstas no serían correctas para el caso de las dos primeras fuentes de variación.

Autores como Zamudio y Alvarado (1996) desarrollaron magistralmente la teoría matricial y aplicaron SAS para analizar tres bases de datos para generar, independientemente, un Anava para los DCA, DBCA y DCL. Ellos centraron su atención en la partición de la variación total que fue registrada en la variable respuesta en dos componentes: la de la unidad experimental (UA) y la del error muestral, para el DCA, en la UE ellos alojaron tratamientos y EE, para el DBCA, además de los dos anteriores, también incluyeron repeticiones, para el DCL, la UA quedó definida por hileras, columnas, tratamientos y EE.

El código para SAS que ellos implementaron en su contribución científica genera una partición de los efectos asociados a la UA y la estimación del error muestral. Con los procedimientos que son considerados en el presente estudio, InfoStat, InfoGen y SAS podrían validar los resultados anteriores y adicionalmente, serían generados los valores para el error conjunto, conformado por el error experimental y el error muestral.

Investigaciones como las de Zamudio y Alvarado (1996) refieren al uso del cociente que se origina entre los cuadrados medios del error experimental sobre el del error muestral como prueba de hipótesis pertinente en los diseños DCA, DBCA y DCL, Martínez (1988) también lo realizó para el DCA y Freund y Wilson (1993) realizaron este procedimiento en un DBCA. En contraparte, Gomez y Gomez (1988) sólo utilizaron ambos errores para estimar sus varianzas correspondientes.

En el presente estudio, las salidas muestran cómo se obtuvo la significancia estadística en las diferentes fuentes de variación del Anava en los tres diseños experimentales de referencia, como lo sugirieron Zamudio y Alvarado (1996); Martínez (1988); asimismo, se muestra la clasificación de las medias de tratamientos usando la prueba de la diferencia mínima significativa honesta (DMSH), también denominada prueba de Tukey. En el Anava se indican los términos de error apropiados para probar las hipótesis estadísticas de interés para los usuarios cuando se aplican InfoStat y InfoGen (véanse las imágenes correspondientes), si en el procedimiento se suprimen éstos en el diseño DBCA se obtendrá el resultado generado por SAS que presentaron Freund y Wilson (1993).

Como puede observarse en las imágenes que se muestran en la presente investigación, se consideró que las variedades de maíz que fueron evaluadas con base en la dehiscencia del polen, son consideradas como factor de efectos fijos. En este contexto, el denominador que origina su valor de F en tratamientos, hileras o columnas siempre es el cuadrado medio del error experimental, como lo sugirieron Martínez (1988); Zamudio y Alvarado (1996); Sahagún (1998); Restrepo (2007a, 2007b).

Sin lugar a dudas, SAS es el paquete estadístico más versátil y más rápido que existe en la actualidad para el diseño y análisis de experimentos, especialmente para aquellos cuya estructura de tratamientos es más compleja, como los que se analizan en series de experimentos en tiempo y espacio, en arreglos de parcelas divididas, subdivididas o bloques divididos, en ensayos 2^n , 3^n , ó 4^n , así como en diferentes tipos de Látices (Martínez, 1988; Sánchez, 1995; SAS, 1998; González *et al.*, 2019), pero su licencia comercial es más costosa que la de InfoStat y InfoGen. Alternativamente, el usuario podría descargar gratuitamente a través de la internet las versiones académicas de prueba de estos tres paquetes, pero es más fácil y rápido hacerlo para InfoStat y InfoGen y estos dos últimos, podrían potenciar su utilidad usando R-Software, que también está disponible gratuitamente a través de la internet.

Conclusiones

Los cálculos manuales que son presentados en este estudio fueron validados correctamente por las versiones académicas de prueba de los tres paquetes estadísticos; InfoStat e InfoGen, deben preferirse sobre SAS debido a que ambos se pueden descargar rápida y fácilmente desde su sitio WEB, pero SAS superó a ambos al generar en menor tiempo, con menor esfuerzo y en un solo procedimiento, un análisis de varianza con submuestreo en los diseños experimentales completamente al azar, bloques completos al azar y cuadro latino, así como en la aplicación de una comparación de medias de tratamientos con la prueba de Tukey. Adicionalmente, las licencias comerciales de InfoStat y InfoGen son más baratas.

Literatura citada

- Balzarini, M. G.; González, L.; Tablada, M.; Casanoves, F.; Rienzo, J. A. y Robledo, C. W. 2008. Manual del usuario de infoStat. Ed. Brujas, Córdoba, Argentina. 348 p.
- Balzarini, M. G. y Di Rienzo, J. A. 2016. InfoGen. FCA. Universidad Nacional de Córdoba, Argentina. <http://www.info-Gen.Com.mx>.
- Di Rienzo, J. A.; Casanoves, F.; Balzarini, M. G.; González, L.; Tablada, M. y Robledo, C. W. 2008. InfoStat. Grupo InfoStat, FCA. Universidad Nacional de Córdoba, Argentina: <https://www.InfoStat.com.ar>.
- Freund, R. J. and Wilson, W. J. 1993. Statistical methods. First Ed. Academic Press, Inc. San Diego, CA. USA. 440-452 pp.
- Gomez, K. A. and Gomez, A. A. 1984. Statistical procedures for agricultural research. 2nd. Ed. John Wiley & Sons, Inc. Printed in singapore. 680 p.
- González, H. A.; Vázquez, G. L.; Sahagún, C. J. y Rodríguez, P. J. E. 2008. Diversidad fenotípica de variedades e híbridos de maíz en el Valle Toluca-Atlacomulco. México. Rev. Fitotec. Mex. 31(1):67-76.
- González, A.; Pérez, D. J.; Sahagún, J.; Franco, O.; Morales, E. J.; Rubí, M.; Gutiérrez, F. y Balbuena, A. 2010. Aplicación y comparación de métodos univariados para evaluar la estabilidad de maíces del Valle Toluca-Atlacomulco. México. Rev. Agron. Costarr. 34(2):129-143.
- González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R., F.; Franco, M.; J. R. P.; 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.
- Hansen, M. J.; Beard, J. T. D. and Hayes, D. B. 2006. Sampling and experimental design. Chapter 3. *In*: analysis and interpretation of freshwater fisheries data. Guy, Ch. S. and. Brown, M. L. Eds. American Fisheries Society. 30-45 pp. Doi: <https://doi.org/10.47866/9781888569773>.
- Jasso, B. G.; González, H. A.; Pérez, L. D. J.; Franco, M. J. R. P.; Rubí, A. M. y Mejía, C. J. 2022. Uso de Opstat para validar resultados en un dialéctico parcial con ocho líneas de maíz evaluadas en un ambiente. Rev. Mex. Cienc. Agríc. 13(1):41-52.
- Martínez, G. A. 1988. Diseños experimentales. Métodos y elementos de teoría. Editorial Trillas, 1^{ra}. Ed. México, DF. 756 p.

- Pérez, L. D.; Jasso, B. G.; Saavedra, G. C.; Franco, M. J. R. P.; Ramírez, D. J. F. y González, H. A. 2022. Uso de artificios en Opstat para analizar series de experimentos en dialélico parcial. *Rev. Mex. Cienc. Agríc.* 13(2):273-287.
- Piepho, H. P.; Büchse, A. and Emrich, K. 2003. A Hitchhiker's guide to mixed models for randomized experiments. *J. Agron. Crop Sci.* 189(3):310-322.
- Restrepo, L. F. 2007. Diagramas de estructuras en el análisis de varianza. *Rev. Colomb. Cienc. Pec.* 20(2):202-208.
- Restrepo, B. L. F. 2007 b. La esperanza del cuadrado medio. *Rev. Colomb. Cienc. Pec.* 20(2):193-201.
- Sahagún, C. J. 1991. Utilidad del análisis de varianza en el estudio de la interacción entre genotipos y ambientes. *Xilonen.* 1(1):21-32.
- Sahagún, C. J. 1998. Construcción y análisis de los modelos fijos aleatorios y mixtos. Universidad Autónoma Chapingo (UACH)-Departamento de Fitotecnia-Programa Nacional de Investigación en Olericultura. Boletín técnico núm. 3. 64 p.
- Sánchez, G. J. J. 1995. El análisis Biplot en clasificación. *Rev. Fitotec. Mex.* 18(2):188-203.
- Statistical Analysis System (SAS). 1998. SAS/STAT Users Guide. Release 6.03 SAS Institute. Carry North Caroline, USA.
- Zamudio, S. F. J. y Alvarado, S. A. A. 1996. Análisis de diseños experimentales con igual número de submuestras. 1^{ra}. Ed. Universidad Autónoma Chapingo (UACH)-División de Ciencias Forestales. México, DF. 85 p.