Revista Fitotecnia Mexicana
ISSN: 0187-7380
revfitotecniamex@gmail.com
Sociedad Mexicana de Fitogenética, A.C.
México

Sahagún Castellanos, Jaime

Evaluación de genotipos en heterogeneidad meteorológica intrarregional. Confusión vs. anidamiento de años en localidades

Revista Fitotecnia Mexicana, vol. 30, núm. 1, enero-marzo, 2007, pp. 97-104
Sociedad Mexicana de Fitogenética, A.C.
Chapingo, México

Ensayo Científico **Rev. Fitotec. Mex. Vol. 30 (1): 97 - 104, 2007**

EVALUACIÓN DE GENOTIPOS EN HETEROGENEIDAD METEOROLÓGICA INTRARREGIONAL. CONFUSIÓN *vs.* ANIDAMIENTO DE AÑOS EN LOCALIDADES

EVALUATION OF GENOTYPES UNDER INTRAREGIONAL METEOROLOGICAL HETEROGENEITY. CONFOUNDING *vs.* NESTING IN YEARS AND LOCATIONS

Jaime Sahagún Castellanos

Programa Universitario de Investigación y Servicio en Olericultura, Depto. de Fitotecnia de la Universidad Autónoma Chapingo. Km 38.5 Carr. México-Texcoco. 56230, Chapingo, Edo de México. Tel. y Fax: 01(595) 952-1642.
* *Autor para correspondencia* (jsahagun@correo.chapingo.mx)

RESUMEN

Con cierta frecuencia la evaluación de genotipos (G) en varias localidades (L) durante varios años (A) se hace en cada uno de los niveles de un factor B en un diseño en bloques al azar. En este trabajo se muestran los análisis de varianza (ANAVA) correspondientes a este tipo de evaluación en cultivos anuales de temporal o secano en dos casos: A anidado en L y A confundido deliberadamente con L. Con base en estos análisis se cumplió el objetivo principal de determinar el efecto que en el nivel de significancia de las pruebas de hipótesis de las fuentes de variación de interés genotécnico y en los estimadores de la varianza de los factores G, B y de su interacción, tiene la confusión deliberada de A con L cuando en realidad A se anida en L. Los estimadores de G, B y BG mostraron sesgos positivos que involucraron la varianza de las interacciones GL, BL y GBL, respectivamente. Similarmente, para las pruebas de hipótesis del ANAVA para G, B y GB el valor del estadístico de la prueba de F aumentó en tanto que el valor que tiene que alcanzar este estadístico para que haya significancia estadística se redujo, lo que ocasionó un aumento en el nivel de significancia asumido. Adicionalmente, sólo en las pruebas de hipótesis de A/L (A anidado en L), B(A/L), G(A/L), BG(A/L) y R/A/L (R es repeticiones) hubo diferencias entre el uso y la ausencia de parcelas divididas con B como factor de parcela grande en modelo aleatorio y mixto (sólo G o B es factor de efectos fijos).

Palabras clave: Componentes de varianza, pruebas de hipótesis, series de experimentos, factor fijo, factor aleatorio.

SUMMARY

Frequently the evaluation of genotypes (G) in several locations (L) during several years (Y) is made under each level of a factor B in a randomized complete block design. In this study the analyses of variance (ANOVA) for the evaluation of rainfed annual crops are shown according to two cases: Y nested in L and Y deliberately confounded with L. Based on these analyses, the main objective was to determine the effect that the use of a model where L and Y are confounded when actually Y is nested in L, has upon the level of statisticcal significance of the F tests and on the estimators of the variances of G, B, and the interaction GB. These estimators showed positive biases involving the variance of the interactions GL, BL, and GBL, respectively. For the tests of the hypotheses for G, B, and GB, the ANOVA F values were greater and the values required to reach statistical significance decreased causing an increase in the assumed level of significance. In addition, only in the tests for Y/L (Y nested in L), B(Y/L), G(Y/L), BG(Y/L), and R/Y/L (R is "replicates"), differences were found between the use and the absence of split plots with B being the whole plot factor for random or mixed (when only G or B is a fixed-effect factor) model.

Index words: Variance components, hypothesis testing, series of experiments, fixed factor, random factor.

INTRODUCCIÓN

El estudio de las series de experimentos desarrolladas para evaluar genotipos (G) en varias localidades (L) durante dos o más años (A) ha sido abordado desde hace tiempo (*v.g.*, Cochran y Cox, 1957; Gómez y Gómez, 1984; Martínez, 1988). En la investigación genotécnica se presentan casos en que además de los efectos de las localidades y años, también interesa los efectos de los niveles de otro factor (B), como densidades de siembra, dosis de fertilización, láminas de riego, etc. Frecuentemente los niveles de los factores B y G se asignan a las parcelas grandes y chicas, respectivamente, de un diseño en bloques al azar con parcelas divididas que se utiliza en el experimento de cada combinación de año y localidad (McIntosh, 1983; Sahagún, 2000).

Típicamente se considera que la relación de cada uno de los cuatro factores (A, L, G y B) con cada uno de los restantes es de cruzamiento (*v.g.*, McIntosh, 1983; Nyquist, 1991). Nyquist (1991) encontró un sesgo en el estimador de la varianza genética producido por ignorar una debida relación de cruzamiento entre A y L y en su lugar confundir completa y deliberadamente A con L, como lo hacen en forma sistemática algunos autores (*v.g.*, Hallauer y Miranda, 1981).

Recibido: 30 de Junio del 2004.
Aceptado: 10 de Octubre del 2006.

Sahagún (2000) considera que hay casos en que dentro de cada año hay variación importante en la manifestación de los agentes meteorológicos (como cantidad y distribución de la lluvia, temperatura, ocurrencia e intensidad de vientos, granizadas y heladas, etc.) en las diferentes localidades, mientras que el tipo de suelo, la altura sobre el nivel del mar, etc. son constantes en cada localidad. Para estas situaciones, afirma este autor, "años" debe estar anidado en "localidades", ya que basar un análisis en un modelo en que "años" se confunde deliberadamente con "localidades" también produce sesgos en la estimación de la varianza de G y, consecuentemente, en el uso que se dé a esta estimación (estimación de heredabilidad, respuesta a la selección, etc).

Los efectos que tiene la indebida confusión de "años" con "localidades" en otros rubros de interés, sin embargo, se desconocen. En particular, el efecto en el nivel de significancia de las pruebas de F de las fuentes de variación más importantes en términos del mejoramiento genético no se ha investigado, a pesar de que suele ser crucial en la toma de decisiones por el fitomejorador, que frecuentemente se basa en la significancia estadística, o en su ausencia, de estas fuentes de variación (con un nivel frecuentemente preestablecido). Tampoco se conocen los cambios que en la estimación de componentes de varianza y en la prueba de hipótesis implica el uso, muy común, de un diseño en bloques al azar sin parcelas divididas para los factores B y G. Esta investigación se hizo con el fin de producir la información que permita conocer el efecto y cambios arriba referidos.

MODELOS Y MÉTODOS DE ANÁLISIS

En este trabajo se utilizó la notación y métodos descritos por Sahagún (2000) en un estudio que abordó problemas similares a los de este trabajo (una de las diferencias es que este autor utilizó parcelas divididas con los niveles de los factores G y B asignados a parcela grande y chica, respectivamente). Por esa razón su descripción se reducirá a lo indispensable.

En lo que en este estudio fue el modelo 1, el factor "años" (A) se anidó en el factor "localidades" (L) y el factor "bloques" (R) se anidó en A. Como se consideró un diseño experimental en bloques al azar con una asignación completamente aleatoria de las *bg* combinaciones de los *b* y *g* niveles de los factores B y G, respectivamente, a las *bg* parcelas de cada bloque, los datos se describieron según el modelo

$$
\begin{aligned}
Y_{ijkmn} = \mu + L_i + A_{j(i)} + R_{k(ij)} + B_m + (BL)_{im} + (AB)_{mj(i)} \\
+ G_n + (GB)_{mn} + (GL)_{in} + (GBL)_{inm} + (GA)_{nj(i)} \\
+ (GBA)_{nmj(i)} + E_{ijkmn} \qquad \text{(Ec. 1)}
\end{aligned}
$$

donde i = 1, 2, …, *l* ; j = 1, 2, …, a ; k = 1, 2, …, r ; m = 1, 2, …, b ; n = 1, 2, …, g ; μ es la media general; L_i es el efecto de la localidad i; $A_{j(i)}$ es el efecto del año j en la localidad i; $R_{k(ij)}$ es el efecto de la repetición k del año j en la localidad i; B_m es el efecto del nivel m del factor B; G_n es el efecto del genotipo n, y E_{ijkmn} es el error correspondiente a Y_{ijkmn}. Los términos restantes del modelo son interacciones entre los efectos de los niveles de los factores A, L, G y B involucrados.

Si los factores A y L se confunden completa y deliberadamente haciendo de cada combinación de año y localidad uno de los *al* niveles de otro factor, aquí representado por C, el modelo es de la forma:

$$
\begin{aligned}
Y_{ijkm} = \mu + C_i + R_{j(i)} + B_k + (CB)_{ik} + G_m + (GC)_{im} \\
+ (GB)_{km} + (GBC)_{ikm} + E_{ijkm} \qquad \text{(Ec. 2)}
\end{aligned}
$$

donde i = 1, 2,…, *l*a ; j = 1, 2,…, r ; k = 1,2, …, b ; m = 1,2, …, g; μ es la media general; C_i es el efecto del ambiente i; $R_{j(i)}$ es el efecto de la repetición j del ambiente i; B_k es el efecto del nivel k del factor B; G_m es el efecto del genotipo m, y E_{ijkm} es el error correspondiente a Y_{ijkm}. Los términos restantes son efectos de interacción entre los efectos de los niveles de los factores que se muestran en cada caso. En ambos modelos (Ecuaciones 1 y 2) se consideró que los términos aleatorios no están correlacionados y que la variable respuesta se distribuye normalmente.

Para determinar las esperanzas de cuadrados medios cuando los modelos se consideraron como aleatorios se usaron las reglas que para tal efecto han sido publicadas (*v. g.*, John, 1971; Searle, 1971; Snedecor y Cochran, 1980). Para los modelos mixtos, la determinación de esperanzas de cuadrados medios se hizo como Lindman (1992) lo describe.

Las pruebas de F del análisis de varianza se basaron en el concepto de diseño insesgado de Kempthorne (1952), y cuando no fue posible la realización de una prueba con base en dos cuadrados medios se recurrió al procedimiento descrito por Satterthwaite (1946). Finalmente, la generación de estimadores de componentes de varianza se basó en el método de momentos.

RESULTADOS Y DISCUSIÓN

Diferencias entre pruebas de hipótesis

Con el modelo 1 (Ec. 1) [donde A se anida en L (A/L) y no hay parcelas divididas], aleatorio o mixto (sólo G tiene efectos fijos), de las pruebas de hipótesis (Cuadros 1 y 2) que involucran a G sólo las de las fuentes de variación BG(A/L) y G(A/L) difieren de las que obtuvo Sahagún (2000) del modelo en que los niveles de los factores B y G se asignan a parcela grande y chica, respectivamente. Otras pruebas que también difieren son las de A/L, R/A/L y B(A/L). Considérese ahora el caso en que A y L se confunden (Ec. 2 y Cuadros 3 y 4); de las pruebas de F de las fuentes de variación que involucran G, sólo la de GBC se hace en forma diferente, con respecto a lo que encontró Sahagún (2000) con base en parcelas divididas. De las pruebas de F restantes también difieren las de C, R/C y CB. En ambos casos (A anidado en L y A confundido con L) las diferencias se deben a que en cada una de esas pruebas interviene al menos un cuadrado medio de error y a que los tres cuadrados medios de error: dos en parcelas divididas (Sahagún, 2000) y en uno en el análisis del modelo 1 aquí estudiado, son diferentes.

Cuando en los modelos 1 y 2 (con A anidado en L y A confundido con L, sin parcelas divididas) sólo el factor G es de efectos fijos, los resultados que se obtengan para G (solo o en interacciones) reflejan a los que se obtendrían para B en el caso en que sólo B fuera de efectos fijos en los mismos modelos. Por esta razón sólo se muestran los análisis de los casos en que el único factor de efectos fijos es G (Cuadros 2 y 4). Los resultados que de estos análisis se generen para G pueden hacerse válidos para el caso en que sólo B es fijo por mera sustitución de G por B y de B por G. Esto es aplicable para lo que resulte para estimación de componentes de varianza y para pruebas de hipótesis.

Estimación de componentes de varianza

En lo que sigue, se determinará si los estimadores de la varianza de los efectos de G (σ^2_G), de B (σ^2_B) y de GB (σ^2_{GB}) que produce el análisis correspondiente al modelo aleatorio basado en la confusión deliberada de A con L (Ec. 2) cuando la relación real es de anidamiento de A en L (Ec. 1) son insesgados. Paralelamente se irá generando información útil para determinar los efectos que la adopción errónea de modelo produce en el nivel de significancia de las pruebas de hipótesis asociada a las fuentes de variación de G, GB y B.

Según la información del Cuadro 3, para el modelo aleatorio en que A está confundido con L (Ec. 2), el estimador de σ^2_G es

$$\hat{\sigma}^2_G = \left[G + GBC - GC - GB\right] / \left(abl r\right) \qquad \text{(Ec. 3)}$$

En la Ec. 3, los términos del paréntesis rectangular representan los cuadrados medios de los factores e interacciones indicados. Para investigar el efecto que tiene el uso indebido del modelo 2 (A y L confundidos entre sí) en lugar del modelo 1 (A anidado en L), la suma de cuadrados (SC) del cuadrado medio GBC (Cuadro 3) se expresará en términos de SCs correspondientes al modelo en que hay anidamiento de A en L (Cuadro 1), que luego se reexpresarán en términos de cuadrados medios. Así, el cuadrado medio GBC es expresable como una combinación lineal de dos cuadrados medios en la forma:

$$GBC = \frac{l-1}{al-1} GBL + \frac{(a-1)l}{al-1}\left[GB(A/L)\right]$$

De este resultado también se obtiene que:

$$GBC = GBL - \frac{l(a-1)}{al-1} GBL + \frac{(a-1)l}{al-1} GB(A/L) \qquad \text{(Ec. 4)}$$

Similarmente, el cuadrado medio GC del modelo 2 (Ec. 2) puede ser expresado en términos de los cuadrados medios GL y G(A/L) del modelo (1) (Ec. 1) en la forma:

$$GC = GL - \frac{l(a-1)}{al-1} GL + \frac{(a-1)l}{al-1} G(A/L) \qquad \text{(Ec. 5)}$$

Con base en las Ecs. 3 y 5, $\hat{\sigma}^2_G$ (Ec. 3) es expresable como

$$\hat{\sigma}^2_G = \left[G + GBL - GL - GB\right] / \left(abl r\right) + \frac{l\left(a-1\right)}{al-1}\left[GL - GBL - G\left(A/L\right) + GB\left(A/L\right)\right] / \left(abl r\right) \qquad \text{(Ec. 6)}$$

Según la Ec. 6 y las expresiones para los estimadores de los componentes de varianza σ^2_G y σ^2_{GL} del modelo aleatorio con anidamiento de A en L (Cuadro 1), el valor esperado (E) de $\hat{\sigma}^2_G$ es:

Cuadro 1. Esperanzas de cuadrados medios y pruebas de F de una serie de experimentos [con r repeticiones (R) cada uno] en a años (A) y l localidades (L) para evaluar g genotipos (G) en b niveles de un factor B. Modelo aleatorio [Ecuación 1].

		Esperanzas de cuadrados medios[1]													
FV	CM	(13)	(12)	(11)	(10)	(9)	(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)	F
L	C_1	p[2]	p	p	p	p			p	p		p	p	p	$[C_1+C_6+C_{10}+C_{11}]/[C_2+C_5+C_9+C_{12}]$
A/L	C_2	p	p	p					p			p	p		$[C_2+C_{12}+C_{13}]/[C_3+C_6+C_{11}]$
R/A/L	C_3	p										p			C_3/C_{13}
B	C_4	p	p		p		p		p	p	p				$[C_4+C_{10}]/[C_5+C_8]$
BL	C_5	p	p		p				p	p					$[C_5+C_{12}]/[C_6+C_{10}]$
B(A/L)	C_6	p	p						p						C_6/C_{12}
G	C_7	p	p	p	p	p	p	p							$[C_7+C_{10}]/[C_8+C_9]$
GB	C_8	p	p		p		p								C_8/C_{10}
GL	C_9	p	p	p	p	p									$[C_9+C_{12}]/[C_{10}+C_{11}]$
GBL	C_{10}	p	p		p										C_{10}/C_{12}
G(A/L)	C_{11}	p	p	p											C_{11}/C_{12}
GB(A/L)	C_{12}	p	p												C_{12}/C_{13}
Error	C_{13}	p													- - - -

[1] Cada número se refiere al producto de un coeficiente por el componente de varianza asociada a la fuente de variación cuyo cuadrado medio tiene tal número. Los coeficientes de los componentes en (13), (12),…,(2), (1) son 1, r, rb, ra, rab, ral, ralb, gr, gra, grab, gb, gbr y gbra, respectivamente.
[2] El símbolo p significa la presencia del componente de varianza que corresponde al número y fuente de variación indicados.
[3] A/L significa que años está anidado en localidades.

$$E\left[\hat{\sigma}^2_G\right] = \sigma^2_G + \frac{l\left(a-1\right)}{al-1} E\left[GL - GBL - G(A/L) + GB\left(A/L\right)\right]/\left(abl r\right)$$

$$= \sigma^2_G + \frac{a-1}{al-1}\sigma^2_{GL} \qquad \text{(Ec. 7)}$$

Así, en la Ec. 7 la confusión de A con L cuando la relación es de anidamiento de A en L produce un estimador $\hat{\sigma}^2_G$ que tiene un sesgo: $\left[\left(a-1\right)/\left(al-1\right)\right]\sigma^2_{GL}$. Este resultado es coincidente con el que debió haber obtenido Sahagún (2000) para parcelas divididas de no haber cometido un error en la derivación de $E\left(\hat{\sigma}^2_G\right)$.

Por analogía, de los resultados para $\hat{\sigma}^2_G$ (Ecs. 6 y 7), para $\hat{\sigma}^2_B$ debe encontrarse que

$$\hat{\sigma}^2_B = \left[B + GBL - BL - GB\right]/ abl r + \frac{l(a-1)}{al-1}\left\{BL - GBL - B(A/L) + GB(A/L)\right\}$$

,y que

$$E\left(\hat{\sigma}^2_B\right) = \sigma^2_B + \left[(a-1)/(al-1)\right]\sigma^2_{BL} \qquad \text{(Ec. 8)}$$

El estimador de σ^2_{GB}, generado en el modelo en que se confunde A con L se expresa en función de dos cuadrados medios (Cuadro 3):

$$\hat{\sigma}^2_{GB} = \left[GB - GBC\right]/\left(alr\right)$$

En términos del modelo con anidamiento de A en L (Ec. 2), de acuerdo con la forma de GBC en la Ec. 4, $\hat{\sigma}^2_{GB}$ también es expresable como

$$\hat{\sigma}^2_{GB} = \left[GB - GBL\right]/\left(alr\right) + \frac{l\left(a-1\right)}{al-1}\left[GBL - GB\left(A/L\right)\right]/\left(alr\right).$$

Además, según los estimadores de σ^2_{GB} y σ^2_{GBL} que producen las esperanzas de cuadrados medios del Cuadro 1:

$$E\left(\hat{\sigma}^2_{GB}\right) = \sigma^2_{GB} + \frac{a-1}{al-1}\sigma^2_{GBL} \qquad \text{(Ec. 9)}$$

La forma y sesgo de los estimadores de σ^2_{GB} y σ^2_G aquí encontrados coinciden con lo que al respecto se encontró con parcelas divididas (Sahagún, 2000). Esto es debido a que en ambos casos los cuadrados medios involucrados son los mismos; lo que no ocurrió en el caso de los estimadores de $\sigma^2_{A/L}$, $\sigma^2_{B(A/L)}$, $\sigma^2_{G(A/L)}$, $\sigma^2_{BG(A/L)}$ y $\sigma^2_{R/A/L}$ [las varianzas de los efectos de A/L, B(A/L), G(A/L), BG(A/L) y R/A/L, respectivamente].

Para el modelo en que sólo B es de efectos fijos y A se ha confundido con L, los estimadores de σ^2_G y σ^2_{GB} expresados en términos del modelo en que A se anida en L deben ser los mismos que los descritos para el modelo aleatorio (Ecs. 7 y 9). Esto es así debido a que cuando sólo B es de efectos fijos, los cuadrados medios (y sus valores esperados) que se involucran en la estimación de σ^2_G y σ^2_{GB}

son los mismos que los del modelo en que todos los efectos son aleatorios (Cuadros 1 y 2).

El sesgo de cada uno de los estimadores de σ^2_G, σ^2_{GB} y σ^2_B anteriormente aludidos (Ecs. 7, 8 y 9) será positivo si $a > 1$ y el componente de varianza involucrado en el sesgo es positivo (no cero). Por otra parte, cuando $a = 1$ (A deja de ser un factor de clasificación) el sesgo es igual a cero, lo que resulta evidente porque con un único factor (L) no hay posibilidad de cometer error respecto a confundir en lugar de anidar factores. Con $l = 1$, en cambio, los sesgos de $\hat{\sigma}^2_G$, $\hat{\sigma}^2_{GB}$ y $\hat{\sigma}^2_B$ serían σ^2_{GL}, σ^2_{BL} y σ^2_{GBL}, pero éstos se confundirían con σ^2_G, σ^2_B y σ^2_{GB}, respectivamente, ya que con una localidad no hay grados de libertad para las fuentes de variación GL, BL y GBL.

Los efectos que puede causar el sesgo de los estimadores de estos tres componentes de varianza en la definición de estrategias de evaluación de genotipos y en estimación de la heredabilidad y la respuesta a la selección, fueron discutidos por Sahagún (2000) y son igualmente válidos en el contexto de la investigación que aquí se hizo.

Cuadro 2. Esperanzas de cuadrados medios y pruebas de F de una serie de experimentos [con r repeticiones (R) cada uno] en l localidades (L) y a años (A) para evaluar g genotipos (G) en b niveles de un factor B. Sólo G es de efectos fijos [Modelo (1)].

FV	CM	Esperanzas de cuadrados medios[1]													F
		(13)	(12)	(11)	(10)	(9)	(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)	
L	C_1	p[2]							p	p		p	p	p	$[C_1+C_6]/[C_2+C_5]$
A/L	C_2	p							p			p	p		$[C_2+C_{13}]/[C_3+C_6]$
R/A/L	C_3	p										p			C_3/C_{13}
B	C_4	p							p	p	p				C_4/C_5
BL	C_5	p							p	p					C_5/C_6
B(A/L)	C_6	p							p						C_6/C_{13}
G	C_7	p	p	p	p	p	p	p							$[C_7+C_{10}]/[C_8+C_9]$
GB	C_8	p	p		p		p								C_8/C_{10}
GL	C_9	p	p	p	p	p									$[C_9+C_{12}]/[C_{10}+C_{11}]$
GLB	C_{10}	p	p		p										C_{10}/C_{12}
G(A/L)	C_{11}	p	p	p											C_{11}/C_{12}
GB(A/L)	C_{12}	p	p												C_{12}/C_{13}
Error	C_{13}	p													- - -

[1] Cada número representa un coeficiente que multiplica al componente de varianza o pseudovarianza asociada a la fuente de variación cuyo cuadrado medio tiene ese número. Los coeficientes de los componentes en (13), (12),…(2), (1) son 1, r, rb, ra, rab, ral, ralb, gr, gra, grab, gbr y gbra, respectivamente.
[2] El símbolo p significa presencia del componente de varianza o pseudovarianza que corresponde al número y FV indicados.
[3] A/L significa que años está anidado en localidades.

Cuadro 3. Esperanzas de cuadrados medios y pruebas de F de una serie de experimentos [con r repeticiones (R) por experimento] en a x l ambientes (C) para evaluar g genotipos (G) en b niveles de un factor B. Todos los factores son aleatorios [Modelo (2)].

FV	CM	Esperanzas de cuadrados medios[1]									F
		E	GBC	GB	GC	G	CB	B	R/C	C	
C	C_1	p[2]	p		p		p		p	p	$[C_1+C_8+C_9]/[C_2+C_4+C_6]$
R/C	C_2	p							p		$C_2//C_9$
B	C_3	p	p	p			p	p			$[C_3+C_8]/[C_4+C_7]$
CB	C_4	p	p				p				C_4/C_8
G	C_5	p	p	p	p	p					$[C_5+C_8]/[C_6+C_7]$
GC	C_6	p	p		p						C_6/C_8
GB	C_7	p	p	p							C_7/C_8
GBC	C_8	p	p								C_8/C_9
E	C_9	p									- - -

[1] Componentes de varianza de los factores o interacciones indicados. Los coeficientes de los componentes de E, GBC, GB, GC,…, R/C, C son 1, r, alr, br, blar, rg, alrg, bg y rbg, respectivamente.
[2] El símbolo p indica presencia del componente en la esperanza del cuadrado medio que corresponda.
[3] R/C significa que "repeticiones" está anidado en ambientes (C).

Nivel de significancia de las pruebas de hipótesis

Sahagún (2000) no estudió las pruebas de hipótesis asociadas con las fuentes de variación del análisis de varianza, en términos del efecto que en el nivel de significancia causa el uso inadecuado del modelo basado en la confusión de A con L, cuando la relación debe ser de anidamiento de A en L; esto se hará a continuación para las fuentes de variación G, GB y B en el modelo aleatorio. Para hacer la prueba de F para el factor G cuando todos los factores son aleatorios y se confunde A con L (Ec. 2), se recurre al cociente de cuadrados medios (Cuadro 3):

$$F_c^G = [G + GBC]/[GC + GB] \qquad \text{(Ec. 10)}$$

De las Ecs. 4 y 5 se encuentra, respectivamente, que:

$$GBC = GBL-[l(a-1)/(al-1)]GBL + GB(A/L) - [l(a-1)/(al-1)]GB(A/L) \qquad \text{(Ec. (11)}$$

$$GC = GL-[l(a1)/(al-1)]\ GL + [l(a-1)/(al-1)]G(A/L) \qquad \text{(Ec.12)}$$

De acuerdo con las Ecs. 11 y 12 para GBC y GC, respectivamente, F_c^G también es expresable como:

$$F_c^G = \frac{G + GBL - \left[l(a-1)/(al-1)\right]GBL + GB(A/L) - \left[(l-1)/(al-1)\right]GB(A/L)}{GB + GL - \left[l(a-1)/(al-1)\right]GL + \left[(a-1)l/(al-1)\right]G(A/L)}$$

$$= \frac{G + GBL - \left[l(a-1)/(al-1)\right]\left[GBL - GB(A/L)\right]}{GB + GL - \left[l(a-1)/(al-1)\right]\left[GL - G(A/L)\right]}$$

En virtud de que $\left[GBL - GB(A/L)\right]/(ar)$ y $\left[GL - G(A/L)\right]/(abr)$ son los estimadores de σ^2_{GBL} y $\sigma^2_{GL} + \sigma^2_{GBL}/b$ en el modelo en que A se anida en L (Cuadro 1), respectivamente, resulta que:

$$F_c^G = \frac{G + GBL - \left[l(a-1)/(al-1)\right]ar\hat{\sigma}^2_{GBL}}{GB + GL - \left[l(a-1)/(al-1)\right]\left[ar\hat{\sigma}^2_{GBL} + abr\hat{\sigma}^2_{GL}\right]} \qquad \text{(Ec. 13)}$$

El cociente de la Ec. 13, en principio de cuadrados medios, debe ser positivo. Por otra parte, la prueba para G en el caso en que A se anida en L y el modelo es aleatorio (Cuadro 1) se basa en el estadístico

$$F_a^G = \frac{G + GBL}{GB + GL} \qquad \text{(Ec. 14)}$$

De acuerdo con las esperanzas de los cuadrados medios de G, GBL, GB y GL (Cuadro 1), debe esperarse que el cociente [G+GBL]/[GB+GL] (Ecuación 14) sea mayor que 1. Además, con estimaciones de componentes de varianza mayores que cero, debe esperarse que el cociente de la Ec. 13 sea mayor que el de [G+GBL]/[GB+GL]. Consecuentemente, tambien debe esperarse que

$$F_c^G > F_a^G \qquad \text{(Ec. (15)}$$

Por analogía, con respecto a la prueba para σ^2_B en el modelo aleatorio en que se confunde A con L, debe esperarse que el estadístico de prueba $\left(F_c^B\right)$ sea mayor que el del modelo con anidamiento $\left(F_a^B\right)$; es decir, debe esperarse que

$$F_c^B > F_a^B \qquad \text{(Ec. 16)}$$

Con respecto a la prueba de F para la interacción GB en el modelo en que se confunde A con L, del Cuadro 3 se obtiene que

$$F_c^{GB} = \frac{GB}{GBC}$$

De acuerdo con la Ec. 4, para el cuadrado medio GBC, resulta que

$$F_c^{GB} = \frac{GB}{GBL - \left[l(a-1)/(al-1)\right]\left[GBL - GB(A/L)\right]}$$

y de aquí que, como $\hat{\sigma}^2_{GBL} = \left[GBL - GB(A/L)\right]/(ar)$ (Cuadro 1), entonces

$$F_c^{GB} = \frac{GB}{GBL - \left[l(a-1)/(al-1)\right]ar\hat{\sigma}^2_{GBL}}$$

Para la prueba de F para σ^2_{GB} en el modelo en que A se anida en L (Cuadro 1) se recurre al cociente

$$F_a^{GB} = \frac{GB}{GBL}$$

Cuadro 4. Esperanzas de cuadrados medios y pruebas de F de una serie de experimentos en a x 1 ambientes (C) para evaluar g genotipos (G) en b niveles de un factor B. Sólo G es de efectos fijos [Modelo (2)] y cada experimento tiene r repeticiones

		Esperanzas de cuadrados medios[1]									
FV	CM	E	GBC	GB	GC	G	CB	B	R/C	C	F
C	C_1	p[2]					p		p	p	$[C_1+C_9]/[C_2+C_4]$
R/C	C_2	p							p		C_2/C_9
B	C_3	p					p	p			C_3/C_4
CB	C_4	p					p				C_4/C_9
G	C_5	p	p	p	P	p					$[C_5+C_8]/[C_6+C_7]$
GC	C_6	p	p		P						C_6/C_8
GB	C_7	p	p	p							C_7/C_8
GBC	C_8	p	p								C_8/C_9
E	C_9	p									- - -

[1] Varianzas o pseudovarianzas de los factores o interacciones indicados. Los coeficientes de E, GBC, GB, GC,…, R/C, C son r, alr, br, blar, rg, alrg, bg y rbg, respectivamente.
[2] p = presencia del componente de varianza o pseudovarianza en la esperanza del cuadrado medio.
[3] R/C significa que repeticiones (R) está anidado en ambientes (C).

De acuerdo con consideraciones análogas a las esgrimidas en la obtención de los resultados de las Ecs. 15 y 16, debe esperarse que

$$F_c^{GB} > F_a^{GB} \qquad \text{(Ec. 17)}$$

Además, debido a que el cociente de cuadrados medios que define a F_c^{GB} (GB/GBC) tiene (con $l > 1$ y $a > 1$) un denominador con más grados de libertad que los del denominador de F_a^{GB} (GB/GBL), el escalar $f^{\alpha}_{(g-1)(b-1),(g-1)(b-1)(l-1)}$ que define el nivel de significancia α preestablecido de la prueba en términos probabilísticos en la forma

$$P\left[F_a^{GB} \geq f^{\alpha}_{(g-1)(b-1),(g-1)(b-1)(l-1)}\right] = \alpha$$

es mayor que el que correspondería a la prueba basada en F_c^{GB}. De acuerdo con esta consideración y con la Ecuación 17, si la prueba de F para GB se basara indebidamente en el modelo 2 (en que se confunde deliberadamente A con L) cuando el modelo 1 (en que A se anida en L) es el que corresponde a la situación real, debe esperarse que la probabilidad de declarar significancia estadística al nivel α que se haya considerado conveniente se incremente.

Considérese ahora la prueba de F del análisis de varianza para el factor G. En virtud de que los estadísticos de prueba F_c^G y F_a^G (Ecs. 10 y 14) difieren sólo en que F_c^G involucra a GBC y GC en el lugar en que F_a^G involucra a GBL y GL, respectivamente, los grados de libertad de F_c^G son más que los F_a^G. Por esta razón y porque $F_c^G > F_a^G$ (Ec. 15), con el uso indebido del modelo con confusión de A con L debe esperarse un incremento injustificado en la probabilidad de declarar significancia estadística del factor G al nivel que, en su caso, se haya determinado. Por analogía, debe esperarse lo mismo con la probabilidad de encontrar significancia estadística en la prueba del factor B que se hace indebidamente con F_c^B en lugar de con F_a^B.

Los resultados obtenidos para las pruebas de F para G, B y GB en el modelo aleatorio también son válidos para las pruebas de G y GB (B y GB) cuando el modelo sea mixto por ser B (G) el único factor de efectos fijos toda vez que los cuadrados medios observados y esperados que intervienen en ellas son iguales en ambos modelos.

Todos los resultados anteriores relacionados con el nivel de significancia de las pruebas de hipótesis pueden influir negativamente en la toma de decisiones. Por ejemplo, en un programa de selección recurrente, con respecto a G su significancia estadística puede hacer que se atribuya, inmerecidamente, a la población que es objeto de mejoramiento genético la característica de contener variabilidad genética que amerite la conducción de un ciclo de selección adicional.

CONCLUSIONES

En la evaluación de genotipos (G) en cada uno de los niveles de un factor B en varias localidades (L) durante dos o más años (A), el uso o la ausencia de parcelas divididas [los niveles del factor adicional B y los genotipos (G) se asignan a parcela grande y chica, respectivamente] no afectan la forma de los estimadores de los componentes de varianza ni de los estadísticos de prueba de las hipótesis asociadas a los factores G y B y a su interacción GB. Sin embargo, la confusión deliberada de A con L cuando A debe estar anidado en L, causó efectos que pueden ser de consideración, en la estimación de componentes de varianza y en la prueba de hipótesis: a) Los estimadores de los componentes de varianza asociados a $G\left(\sigma^2_G\right)$, $GB\left(\sigma^2_{GB}\right)$ y $B\left(\sigma^2_B\right)$ mostraron un sesgo positivo, y b) La probabilidad de encontrar significancia estadística de las fuentes de variación G, B y GB a un nivel preestablecido por el fitomejorador sufrió un indebido incremento. En lo que atañe a los factores G y B e interacción GB que resulten aleatorios cuando G o B sea el único factor de efectos fijos, los estimadores y el nivel de significancia de las pruebas de hipótesis, tuvieron el comportamiento que justamente se describió para el modelo aleatorio.

BIBLIOGRAFÍA

Cochran W G, G M Cox (1973) Experimental Designs. Wiley. New York. 661 p.

Gómez K A, A A Gómez (1984) Statistical Procedures for Agricultural Research. 2nd ed. Wiley. New York. 680 p.

Hallauer A R, J B Miranda Fo (1981) Quantitative Genetics in Maize Breeding. Iowa State University Press. Ames. IA. 468 p.

John P W M (1971) Statistical Design and Analysis of Experiments. The McMillan Company, New York. 356 p.

Kempthorne O (1952) The Design and Analysis of Experiments. Wiley, New York. 631 p.

Lindman H R (1992) Analysis of Variance in Experimental Designs. Springer Verlag. 531 p.

Martínez G A (1988) Diseños Experimentales. Métodos y Elementos de Teoría. Trillas. México. 756 p.

McIntosh M S (1983) Analysis of combined experiments. Agron. J. 75:153-155.

Nyquist W E (1991) Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10: 235-322.

Sahagún C J (2000) Evaluación de genotipos con varios niveles de un factor en series de experimentos. Agrociencia 34:193-206.

Satterthwaite F R (1946) An approximate distribution of estimates of variance components. Biometrics Bull. 2:110-112.

Searle S R (1971) Linear Models. Wiley. New York. 532 p.

Snedecor G W, W G Cochran (1980) Statistical Methods. 7th ed. Iowa State University Press. Ames, IA. 597 p.