

EVALUACIÓN GENOTÍPICA CON CRUZAMIENTO CUANDO HAY ANIDAMIENTO DE AÑOS Y LOCALIDADES

GENOTYPIC EVALUATION WITH CROSSING WHEN THERE IS NESTING BETWEEN YEARS AND LOCATIONS

Jaime Sahagún Castellanos¹

RESUMEN

La importancia de la correspondencia entre modelo y situación experimental que aquél describe ha sido reconocida desde el origen de los principios y teoría básica del modelo lineal general. En la evaluación de genotipos (G) es recomendable hacer con ellos un experimento en cada una de las $a \times l$ combinaciones de a años (A) y l localidades (L). Un modelo que se podría utilizar para esta serie de experimentos es el basado en el anidamiento de A en L [modelo (3)]; otro, frecuentemente utilizado, es el basado en el cruzamiento de A con L [modelo (2)]. Suponiendo que el modelo (3) es el que corresponde a la situación experimental, en este estudio se investigó el efecto del uso del modelo (2) en la evaluación genotípica. El estudio se centró básicamente en la estimación de componentes de varianza de interés genotécnico. Respecto a la varianza genética, su estimador resultó ser insesgado cuando la varianza de la interacción AxG es la misma en cada localidad. El estimador de la varianza de la interacción GxL siempre fue insesgado. En cambio, al estimador de la varianza de la interacción general AxG se le encontró un valor esperado igual a cero. En relación con las varianzas de G y de GxL, presentes en ambos modelos, los estimadores del modelo (2) que, como los del modelo (3), fueron insesgados, tuvieron las varianzas más elevadas. Estas mayores magnitudes de las varianzas de los estimadores del modelo (2) deben incrementar la probabilidad de que se tome decisiones equivocadas en la planeación o ejecución de los programas de mejoramiento genético.

PALABRAS CLAVE ADICIONALES

Precisión, heredabilidad, componentes de varianza, estimadores insesgados, mejoramiento genético.

¹ Departamento de Fitotecnia. Universidad Autónoma Chapingo, Méx. Tel. y Fax: 01(595) 2-1642

SUMMARY

The importance of the relationship between model and experimental situation which is to be described has been recognized since the origin of the principles and basic theory of the general linear model. In the genotypic evaluation it is advisable to make an experiment in each combination of a years (A) and l locations (L). A model that could be used for this series of experiments is the one based on the consideration that A is nested by L [model (3)]. Another model, frequently used, is based on the consideration that A is crossed with L [model(2)]. In this study, the effect of using model (2) when A is nested by L is studied. The study was mainly devoted to the estimation of the variance components for breeding purposes. With respect to the genetic variance its estimator was unbiased when the variance of the AxG interaction was the same in each location. The estimator of the variance of the GxL interaction was always unbiased, and the estimator of the general interaction AxG showed an expected value equal to zero. Model (2) estimators of the effects G and GxL, were; as those from model (3), unbiased, but their variances were larger. These larger variances must increase the probability of making wrong decisions in the planning or execution of plant breeding programs.

ADDITIONAL INDEX WORDS

Precision, heritability, variance components, unbiased estimators, genetic improvement.

INTRODUCCIÓN

Diversos autores (Addelman, 1970; Anderson y McLean, 1974; Searle, 1971, 1987; Wilk y Kempthorne, 1955; entre otros) se han referido a la importancia que tiene el modelo estadístico

que se usa en la descripción de la composición de los valores de las variables que se generan en las unidades experimentales. Es común que para una misma situación experimental se encuentre que la adopción de un modelo lleve a resultados que están más ligados a la naturaleza intrínseca de éste que a la propia situación experimental (Sahagún, 1994). Esto, que es de observancia en las ciencias experimentales en general, tiene características muy peculiares en el contexto de la evaluación de genotipos ligada al mejoramiento genético de los cultivares.

Para una evaluación más objetiva y precisa de un conjunto de genotipos es necesario hacer un experimento, cuyos tratamientos sean esos genotipos, en cada una de varias localidades durante dos o más años. Las repeticiones en cada experimento y la realización de varios experimentos son indispensables para aumentar la precisión y la habilidad para estudiar la interacción entre genotipos y ambientes, respectivamente. Con relación a los modelos asociados a estas series de experimentos, también se ha estudiado la relación entre modelo y situación experimental.

Sahagún (1993) se refirió a cuatro modelos que son utilizados para series de experimentos del tipo descrito. Las diferencias entre tales modelos tienen su origen básicamente en la forma en que se relaciona a los factores años (A) y localidades (L) y a la consideración de si los bloques (B) se anidan sólo en localidades y/o en años: 1) Los factores A y L se consideran cruzados y el factor B anidado sólo en L; 2) El factor A y el L se consideran cruzados y B anidado en ambos; 3) El factor A se considera anidado en el L y B se anida en ambos; 4) Cada combinación de año y localidad se considera como un nivel de un factor denominado ambientes. En su estudio, Sahagún (1994) se refirió en particular a los modelos (2) y (3) y encontró que cuando debiéndose usar el (2) se usa el (3), los estimadores de los componentes

de varianza genética y de interacción entre genotipos y localidades producen un sesgo negativo y positivo, respectivamente. Este autor también discutió el efecto que estos sesgos tienen en la heredabilidad y la respuesta a la selección. Sin embargo, en ese estudio no se hizo referencia a la estimación del componente de varianza de la interacción que involucra a los tres factores, ni a las características de la estimación que se hace con el modelo (2) cuando el (3) es el que corresponde a la situación experimental.

El modelo (2) en el que los factores años y localidades se consideran cruzados, para ser realista requiere que los agentes meteorológicos asociados a los años (cantidad y distribución de precipitación, fecha de ocurrencia de primeras y últimas heladas en un invierno, ocurrencia de granizadas, etc.) incidan de la misma manera en las diferentes localidades del área que es objeto de un programa de mejoramiento. Quizá en algunas regiones agrícolas este requerimiento se satisfaga en un grado aceptable. Sin embargo, la gran variación en frecuencia e intensidad con que se manifiestan estos agentes meteorológicos en localidades relativamente cercanas de algunas regiones sugiere, para esos casos, el uso del modelo (3) en el que los años se anidan en las localidades. Este anidamiento implica que en todas las combinaciones de año y localidad la frecuencia, distribución e intensidad de los agentes meteorológicos en cuestión no se manifiestan de igual manera ni dentro ni entre los años. Considerando los sesgos obtenidos por Sahagún (1994) al utilizar un modelo erróneo, la hipótesis de que el uso del modelo de factores cruzados cuando hay anidamiento conduce a sesgos y bajas precisiones, tiene un significado que amerita estudiar su validez.

Fue en el contexto de la situación anterior en donde se originó el proyecto que dio lugar a este trabajo de investigación dirigido al estudio de las consecuencias del análisis basado en el

modelo que considera el cruzamiento de años y localidades cuando el modelo más realista es el que da cuenta de un anidamiento de años en localidades. Los resultados de este análisis se expresarán en términos de las implicaciones que éstos tengan en el contexto de la calidad de la estimación de componentes de varianza de interés genotécnico.

MÉTODOS Y MODELOS

En este estudio se considerará la evaluación de g genotipos (G) en la zona objeto de un programa de mejoramiento genético, haciendo un ensayo en cada combinación de a años (A) con l localidades (L). Tal evaluación se realizará en un experimento de campo en un diseño en bloques completos al azar con r repeticiones. Se supondrá que la especie corresponde a un cultivo anual en condiciones de temporal [maíz (*Zea mays* L.), frijol (*Phaseolus vulgaris* L.), etc.]. Con base en la forma general del análisis de varianza del modelo (2), usando el método de momentos se obtendrá los estimadores de los componentes de varianza de interés genotécnico. Después, estos estimadores serán analizados suponiendo que el modelo (3) es el correcto, para determinar, desde esta perspectiva, sus características (valor esperado y varianza), mismas que se compararán con las de los estimadores correspondientes generados en el propio modelo (3).

La determinación de las varianzas de los estimadores se basará en el supuesto que establece que, si la variable respuesta sigue una distribución normal, un cuadrado medio CM con gl grados de libertad tiene una varianza igual a $2[E(CM)]^2/gl$ (Searle, 1971).

Cuando los años y las localidades son factores cruzados, un modelo que frecuentemente se usa para describir el valor fenotípico Y_{ijkm} es el que explica este valor en términos de la suma de los

efectos de: una media general (μ), el genotipo i (G_i), la localidad k (L_k), el año m (A_m), el bloque j anidado en la localidad k [$B_{j(k)}$], las interacciones (GL) $_{ik}$, (AG) $_{im}$, (AGL) $_{ikm}$ y el error E_{ijkm} ; es decir,

$$Y_{ijkm} = \mu + G_i + L_k + A_m + B_{j(k)} + (GL)_{ik} + (AG)_{im} + (AL)_{km} + (AGL)_{ikm} + E_{ijkm} \quad (1)$$

El modelo (1), por considerarse que los bloques se anidan sólo en las localidades, es adecuado para especies perennes, o para cultivos anuales cuando los bloques son los mismos en cada localidad durante los a años. En este caso, los factores L , A y B son normalmente considerados como aleatorios en tanto que G puede ser aleatorio o fijo, dependiendo de si la elección de niveles fue al azar o no, respectivamente.

En el caso típico de cultivos anuales, en que el lote experimental dentro de cada localidad es diferente de un año a otro, los bloques se anidan tanto en localidades como en años. Para este caso, un modelo puede ser de la forma,

$$Y_{ijkm} = \mu + G_i + L_k + A_m + B_{j(km)} + (GL)_{ik} + (AG)_{im} + (AL)_{km} + (AGL)_{ikm} + E_{ijkm} \quad (2)$$

En este modelo, exceptuando a μ , todos los términos del lado derecho de la igualdad serán considerados como variables aleatorias normal e independientemente distribuidas con media igual a cero. Además, el nivel p -ésimo (X_p) de cualquier efecto o interacción aleatoria X tendrá una varianza $V(X)$.

Otro modelo expresa que los agentes ambientales asociados con los años (cantidad y distribución de precipitación, patrón de tempe-

Cuadro 1. Información básica del modelo (2) para la estimación y prueba de hipótesis sobre componentes de varianza.

Fuente de variación	Grados de libertad	Cuadrados medios	Esperanza de cuadrados medios									
			σ^2	$r\sigma_{agl}^2$	$ra\sigma_{gl}^2$	$rl\sigma_{ag}^2$	$ral\sigma_g^2$	$g\sigma_{b(al)}^2$	$rg\sigma_{al}^2$	$rag\sigma_l^2$	$rlg\sigma_a^2$	
Años (A)	a-1	A	p	p		p			p	p		p
Localidades (L)	l-1	L	p	p	p				p	p	p	
AL	(a-1)(l-1)	AL	p	p					p	p		
Bloques/AL	(r-1)al	B/AL	p						p			
Genotipos (G)	g-1	G	p	p	p	p	p					
AG	(g-1)(a-1)	AG	p	p		p						
GL	(g-1)(l-1)	GL	p	p	p							
AGL	(g-1)(a-1)(l-1)	AGL	p	p								
Error	(r-1)al(g-1)	E _r	p									

p: presencia del término.

142

Cuadro 2. Información básica del modelo (3) para la estimación y prueba de hipótesis sobre componentes de varianza.

Fuente de variación	Grados de libertad	Cuadrados medios	Esperanza de cuadrados medios							
			σ^2	$r\sigma_{g(a/l)}^2$	$ra\sigma_{gl}^2$	$ral\sigma_g^2$	$g\sigma_{b(a/l)}^2$	$rgl\sigma_a^2$	$rag\sigma_l^2$	
Localidades (L)	(l-1)	L	p	p			p	p		p
Años (A)/L	(a-1)l	A/L	p	p			p	p		
Bloques (B)/A/L	(r-1)al	B/A/L	p				p			
Genotipos (G)	g-1	G	p	p	p	p				
GL	(g-1)(l-1)	GL	p	p	p					
G(A/L)	(g-1)(a-1)l	G(A/L)	p	p						
Error	(r-1)al(g-1)	E _r	p							

p: presencia del término.

raturas, vientos, granizadas, heladas, etc.) no se manifiestan de igual manera en las diferentes localidades dentro de un año ni de un año a otro, en tanto que los rasgos básicos de cada localidad (tipo de suelo, altura sobre el nivel del mar, etc.) son constantes a través del tiempo. Un modelo para esta situación experimental es de la forma

$$Y_{ijkm} = \mu + G_i + (B)_{j(km)} + A_{k(m)} + (AG)_{ik(m)} + L_m + (GL)_{im} + \epsilon_{ijkm} \quad (3)$$

Como en el modelo anterior, exceptuando μ , todos los términos del lado derecho del modelo son variables aleatorias independientes entre y dentro de si. Además, el efecto del nivel p-ésimo (Y_p) de cualquier factor o interacción aleatoria Y tendrá una varianza $V(Y)$.

Como ya se mencionó, para obtener los estimadores de componentes de varianza se utilizará el método de momentos que consiste en establecer un sistema de ecuaciones entre los cuadrados medios y sus correspondientes valores esperados. Considerando como incógnitas a los componentes de varianza, la solución al sistema produce los estimadores de tales componentes. Las esperanzas de cuadrados medios serán determinadas según las reglas que describen An-

derson y McLean (1974), Snedecor y Cochran (1980) y Lindman (1992). Cuando éstas no sean aplicables se recurrirá a la derivación directa de los valores esperados.

RESULTADOS Y DISCUSIÓN

Para los modelos (2) y (3), los Cuadros 1 y 2, respectivamente, muestran las fuentes de variación, los grados de libertad y los cuadrados medios experimentales y esperados.

Los estimadores de componentes de varianza de los modelos (2) y (3) se describen en el Cuadro 3 (Sahagún, 1994). En este cuadro se exhiben las diferencias entre los estimadores correspondientes generados en los dos modelos. Con base en el modelo (2), el estimador de la varianza genética es

$$\sigma_g^2 = [G - GL - AG + AGL] / lar = [G - GL] / lar - [AG - AGL] / lar \quad (4)$$

Además, usando $SC(.)$ para expresar suma de cuadrados, para el cuadrado medio $G(A/L)$ del modelo (3) resulta que

Cuadro 3. Estimadores de componentes de varianza según los modelos (2) y (3).

Componentes de varianza	Estimador	
	Modelo (2)	Modelo (3)
σ_{agl}^2	$[AGL - Er] / r$	- - - -
σ_{gl}^2	$[GL - AGL] / ar$	$[GL - G(A/L)] / ar$
σ_{ag}^2	$[AG - AGL] / r$	- - - - -
$\sigma_{g(a/l)}^2$	- - - -	$[G(A/L) - Er] / r$
σ_g^2	$[G - GL - AG + AGL] / lar$	$[G - GL] / lar$

$$G(A/L) = \frac{SC[G(A/L)]}{(a-1)(g-1)}$$

$$= \frac{SC(AGL) + SC(AG)}{(a-1)(g-1)}$$

$$= \frac{l-1}{l} AGL + \frac{1}{l} AG;$$

es decir,

$$AGL = \frac{l}{l-1} G(A/L) - \frac{1}{l-1} AG.$$

Según esta información y los resultados del Cuadro 3, cuando el modelo (3) es el que mejor refleja a la realidad pero se usa el (2), éste produce un estimador de σ_g^2 cuyo valor esperado es

$$E[\sigma_{g2}^2] = \sigma_g^2 - E\left\{AG - \frac{l}{l-1} G(A/L) + \frac{1}{l-1} AG\right\} / (ra)$$

$$= \sigma_g^2 - E\left\{\frac{l}{l-1} [AG - G(A/L)]\right\} / (ra); \tag{5}$$

es decir, el estimador de σ_g^2 del modelo (2) cuando el modelo (3) es el que debe usarse produce un sesgo de $-E\{[AG - G(A/L)] / [ra(l-1)]\}$.

Otra varianza presente en ambos modelos es σ_{gl}^2 . Con el modelo (2), $\sigma_{gl}^2 = [GL - AGL] / ra$. Para este estimador,

$$\sigma_{gl}^2 = \left[GL - \frac{SC(AGL)}{(g-1)(a-1)(l-1)}\right] / (ra)$$

$$= \left[GL - \frac{SC[G(A/L)] - SC(AG)}{(a-1)(l-1)(g-1)}\right] / (ra)$$

$$= \left[GL - \frac{l[G(A/L)]}{l-1} + \frac{AG}{l-1}\right] / (ra)$$

$$= [GL - G(A/L)] / (ra) + [AG - G(A/L)] / [(l-1)ra], \tag{6}$$

de manera que, de acuerdo con los resultados respectivos del Cuadro 3,

$$E[\sigma_{gl}^2] = \sigma_{gl}^2 + E[AG - G(A/L)] / [(l-1)ra]; \tag{7}$$

es decir, la aplicación del modelo (2) cuando se debe usar el (3) produce un sesgo igual a $E[AG - G(A/L)] / [(l-1)ra]$ en la estimación de σ_{gl}^2 .

En el Cuadro 3 se observa que según el modelo (2) σ_{ag}^2 y σ_{ag}^2 también son estimables. Sin embargo, si el modelo (3) es el correcto, ... ¿Qué se estima con los estimadores $\sigma_{ag/2}^2$ y σ_{ag2}^2 generados en el modelo (2)? Respecto a $\sigma_{ag/2}^2$ se tiene que

$$\sigma_{ag/2}^2 = [AGL - Er] / r$$

$$= \left[\frac{SC(AGL)}{(a-1)(l-1)(g-1)} - Er\right] / r$$

$$= \left[\frac{SC[G(A/L)] - SC(AG)}{(a-1)(l-1)(g-1)} - Er\right] / r$$

$$= [G(A/L) - Er] / r + \frac{1}{(l-1)} [G(A/L) - AG] / r. \tag{8}$$

Por lo tanto,

$$E\{\sigma_{ag/2}^2\} = \sigma_{ag/2}^2 - E\left\{\frac{1}{l-1} [AG - G(A/L)]\right\} / r. \tag{9}$$

Finalmente, respecto al estimador σ_{ag2}^2 del modelo (2), siguiendo un procedimiento similar al anterior se obtendrá que, cuando el modelo (3) es el que debe usarse,

$$\begin{aligned} \sigma_{ag2}^2 &= [AG - AGL]/(lr) \\ &= \left[AG - \frac{G(A/L)}{l-1} + \frac{AG}{l-1} \right] / (lr) \\ &= \frac{l}{l-1} [AG - G(A/L)] / (lr) \\ &= \frac{1}{l-1} [AG - G(A/L)] / r. \end{aligned} \tag{10}$$

Esto es, según (9), σ_{ag2}^2 es un estimador insesgado del sesgo de $\sigma_{ag/2}^2$, visualizado éste como estimador de $\sigma_{g(a/l)}^2$.

La estimación de σ_g^2 es de particular importancia en genotecnia. Cuando el modelo (3) es el que se debe usar y en su lugar se usa el (2) se genera un sesgo de $-E\{[AG - G(A/L)]/(l-1)\}/(alr)$. Por ser σ_g^2 el numerador del cociente correspondiente a la fórmula de heredabilidad, la determinación de la magnitud del sesgo, en su caso, con que se estima σ_g^2 es de importancia singular. Para determinar la magnitud del sesgo se usará la notación en donde un punto en el lugar de un subíndice indicará que se ha sumado sobre tal subíndice, y una raya sobre una variable indicará que se ha promediado sobre los subíndices sumados. Así, respecto a la fuente de variación AG su suma de cuadrados según el modelo (2) es expresable como

$$SC(AG) = r \sum_i \sum_k [\bar{Y}_{i.k} - \bar{Y}_{i..} - \bar{Y}_{.k} + \bar{Y}_{...}]^2$$

En términos del modelo (3), ésta sería expresable como

$$SC(AG) = r \sum_i \sum_k \left[(\bar{AG})_{ik(.)} - (\bar{AG})_{i.(.)} - (\bar{AG})_{.k(.)} + (\bar{AG})_{..(.)} + \varepsilon_{i.k} - \varepsilon_{i..} - \varepsilon_{.k} + \varepsilon_{...} \right]^2 \tag{11}$$

por lo que el valor esperado del cuadrado medio AG es

$$\begin{aligned} E[AG] &= r \sum_i \sum_k \left[\left(\frac{1}{l} + \frac{1}{a} + \frac{1}{r} + \frac{1}{g/a} - \frac{2}{la} - \frac{2}{lg} \right) \sigma_{g(a/l)}^2 \right. \\ &\quad \left. + \left(\frac{1}{rl} + \frac{1}{ral} + \frac{1}{rgl} + \frac{1}{glar} - \frac{2}{rla} - \frac{2}{grl} \right) \sigma^2 \right] / [(a-1)(g-1)] \\ &= \left[(gra + rg + ra + r - 2gr - 2ra) \sigma_{g(a/l)}^2 + (ga + g + a + 1 - 2g - 2a) \sigma^2 \right] / [(a-1)(g-1)] \\ &= \left\{ r(g-1)(a-1) \right\} \sigma_{g(a/l)}^2 + (g-1)(a-1) \sigma^2 \Big/ [(a-1)(g-1)] \\ &= r \sigma_{g(a/l)}^2 + \sigma^2. \end{aligned} \tag{12}$$

En el desarrollo anterior se ha supuesto que cada término $(AG)_{ik(m)}$ tiene una misma varianza ($\sigma_{g(a/l)}^2$). Sin embargo, en una situación en que los factores ambientales de cada año no inciden de la misma manera en cada localidad, la varianza de la interacción puede no ser la misma en todas las localidades. De acuerdo con este razonamiento, si a la varianza del efecto de interacción entre el genotipo i y el año k en la localidad m se le representa por $\sigma_{ag(m)}^2$ entonces, partiendo de (11) se puede encontrar que

$$E[AG] = \sigma^2 + r \sum_m \sigma_{ag(m)}^2 / l. \tag{13}$$

Ahora, respecto a $G(A/L)$, en virtud de que

$$\begin{aligned} SC[G(A/L)] &= r \sum_i \sum_k \sum_m [\bar{Y}_{i.km} - \bar{Y}_{i..m} - \bar{Y}_{.km} + \bar{Y}_{...m}]^2 \\ &= r \sum_i \sum_k \sum_m \left[(AG)_{ik(m)} - (\bar{AG})_{i.(m)} - (\bar{AG})_{.k(m)} + (\bar{AG})_{..(m)} + \varepsilon_{i.km} - \varepsilon_{i..m} - \varepsilon_{.km} + \varepsilon_{...m} \right]^2, \end{aligned} \tag{14}$$

entonces si para cualquier combinación de i y k(m) la varianza del efecto de interacción es la misma, $[\sigma_{g(a/l)}^2]$, de acuerdo con el modelo (3) y lo mostrado en el Cuadro 2,

$$E[G(A/L)] = \sigma^2 + r\sigma_{g(a/l)}^2 \tag{15}$$

Por el otro lado, admitiendo que la varianza de la interacción entre genotipos y años anidados en las localidades puede no ser la misma en todas las localidades, la varianza entre genotipo *i* y el año *k* en la localidad *m* se representará, como ya fue hecho, por $\sigma_{ag(m)}^2$. Así, de acuerdo con (14),

$$\begin{aligned} E[G(A/L)] &= r \sum_i \sum_k \sum_m \left[\left(1 + \frac{1}{a} + \frac{1}{g} + \frac{1}{ag} - \frac{2}{a} - \frac{2}{g} \right) \sigma_{ag(m)}^2 \right. \\ &\quad \left. + \left(\frac{1}{r} \right) \left(1 + \frac{1}{a} + \frac{1}{g} + \frac{1}{ag} - \frac{2}{a} - \frac{2}{g} \right) \sigma^2 \right] / [(a-1)(g-1)] \\ &= \sum_m \left[r(ga + g + a + 1 - 2g - 2a) \sigma_{ag(m)}^2 + \right. \\ &\quad \left. (ga + g + a + 1 - 2g - 2a) \sigma^2 \right] / [(a-1)(g-1)] \\ &= r(a-1)(g-1) \sum_m \sigma_{ag(m)}^2 / [(a-1)(g-1)] + \sigma^2 \\ &= r \sum_m \sigma_{ag(m)}^2 / l + \sigma^2 \tag{16} \end{aligned}$$

De acuerdo con los resultados (12) y (15), cuando se asigna una misma varianza al efecto de la interacción entre genotipos y años independientemente de la localidad de que se trate se tendrá que, partiendo de (5), el estimador de σ_g^2 generado en el modelo (2) cuando el modelo (3) es el que corresponde a la situación experimental, es insesgado. Véase:

$$\begin{aligned} E[\hat{\sigma}_{g2}^2] &= \sigma_g^2 - E\{[AG - G(A/L)] / (l-1)\} / (alr) \\ &= \sigma_g^2 - \{[E(AG) - E[G(A/L)]] / (l-1)\} / (alr) \\ &= \sigma_g^2 \tag{17} \end{aligned}$$

Similarmente, de acuerdo con los resultados (13) y (16), cuando la varianza del efecto de interacción entre genotipos y años (anidados en el factor *L*) puede no ser la misma en todas las

localidades, el estimador de σ_g^2 generado en el modelo (2) aplicado a una situación en que el modelo (3) es un mejor descriptor del valor observado en cada unidad experimental, partiendo de (5), tiene el valor esperado que a continuación se deriva:

$$\begin{aligned} E[\hat{\sigma}_{g2}^2] &= \sigma_g^2 - E\{[AG - G(A/L)] / (l-1)\} / (alr) \\ &= \sigma_g^2 - \{[E(AG)] - E[G(A/L)] / (l-1)\} / (alr) \\ &= \sigma_g^2 \tag{18} \end{aligned}$$

Respecto al estimador de σ_{gl}^2 se encontró que el sesgo de σ_{gl}^2 es $E[AG - G(A/L)] / [(l-1)ar]$. Sin embargo, de acuerdo con la discusión previa, basada en los resultados (13) y (16), este sesgo es igual a cero. Además, $\sigma_{ag/2}^2$ y σ_{ag2}^2 , según las expresiones (9) y (10), respectivamente, son estimadores insesgados de $\sigma_{g(a/l)}^2$ y 0, respectivamente.

Las características del modelo (3) permiten explicar por qué el sesgo $[AG - G(A/L)] / (l-1)$ tiene un valor esperado igual a cero. Si, por ejemplo, la varianza de $AG(\sigma_{ag}^2)$ fuera la misma en cada localidad, con la información experimental de cada localidad se puede estructurar un estimador de σ_{ag}^2 en la misma fórmula del cuadrado medio de $G(A/L)$. Como éste tiene *l* veces más grados de libertad que el cuadrado medio de AG y como cada esperanza de cuadrado medio contiene a σ^2 , debe suceder que $E[G(A/L)] = E[AG]$ y que, en consecuencia, el sesgo sea igual a cero.

En la prueba de la hipótesis $H_0: \sigma_g^2 = 0$ también se refleja el efecto negativo del uso de un modelo que no corresponde a la realidad. Con base en el modelo (2) la prueba se basa en el co-

ciente $[G+AGL]/[AG+AL]$ que se distribuye aproximadamente (Satterthwaite, 1946) como $F_{p,q}$ en donde

$$p = \frac{[G + AGL]^2}{\frac{G^2}{g-1} + \frac{(AGL)^2}{(a-1)(g-1)(l-1)}}, \quad q = \frac{[AG + AL]^2}{\frac{(AG)^2}{(a-1)(g-1)} + \frac{(AL)^2}{(a-1)(l-1)}}$$

Si el modelo (3) es el que corresponde a la situación experimental, de acuerdo con los resultados (5), (7), (9), (10), (13) y (15), sucede que, como en el modelo (2),

$$E[G + ALG] = ra/\sigma_g^2 + E[AG + GL].$$

Sin embargo, aun cuando se verifica esta propiedad deseable de esta prueba (Searle, 1971) el uso del modelo (2) asigna como grados de libertad a p y q en lugar de los que corresponden al modelo (3) que son $g-1$ y $(g-1)(a-1)l$, respectivamente, de acuerdo con el cociente $G/[G(A/L)]$ de la prueba exacta del modelo (3) que es el que hipotéticamente representa la situación experimental. Esta diferencia en grados de libertad debe alterar la probabilidad de rechazar, o no rechazar la hipótesis nula para un nivel dado de significancia. Esto a su vez promoverá la toma de una decisión equivocada por parte del fitomejorador.

En relación con los componentes de varianza de AG y AGL del modelo (2), que no existen en el modelo (3), el uso de aquél cuando éste es el que corresponde a la situación experimental genera discusión y conclusiones sobre parámetros que no son los que suponen ser y que, probablemente, inducen cursos de acción equivocados.

De los estimadores insesgados de σ_g^2 y $\sigma_{g/l}^2$ generados en el modelo (2), respecto a los generados en el modelo (3), que también son insesgados ¿Cuáles son mejores?. Para contestar esta

interrogante se adoptará el criterio según el que un estimador es mejor que otro si, siendo ambos insesgados, su varianza es la menor. Las varianzas de los estimadores de σ_g^2 se describen a continuación. Según (4),

$$v[\sigma_{g2}^2] = v\{[G - GL - AG + AGL]/(lar)\} \\ = \{v[G] + v[GL] + v[AG] + v[AGL]\}/(lar)^2, \quad (19)$$

y

$$v[\sigma_{g3}^2] = v\{[G - GL]/(lar)\} \\ = \{v[G] + v[GL]\}/(lar)^2. \quad (20)$$

Según los resultados (19) y (20), cuando los años están anidados en localidades, el estimador de σ_g^2 que se genera en el modelo (3) tiene una varianza menor que la del que se genera en el modelo (2), en el que se da cuenta de una relación de cruzamiento entre los factores años y localidades.

Respecto a $\sigma_{g/l}^2$, a continuación se determinará la varianza del estimador que se genera en el modelo (2) que, según el Cuadro 3, es de la forma

$$\sigma_{g/l2}^2 = [GL - AGL] / (ar).$$

Considerando que el cuadrado medio AGL se puede escribir como $l\{G(A/L) - AG\} / (l-1)$, respecto a la varianza de $\sigma_{g/l2}^2$ se obtiene que

$$(ar)^2 v[\sigma_{g/l2}^2] = v\{GL - lG(A/L)/(l-1) + AG/(l-1)\}.$$

Los términos segundo y tercero del lado derecho de esta expresión tienen una covarianza igual a $\sigma_{ag/l}^2$. Con esta consideración,

$$(ar)^2 v[\sigma_{ag2}^2] = v[GL] + l^2 v[G(A/L)] / (l-1)^2 - v(AG) / (l-1)^2$$

$$= v[GL] + v[G(A/L)] + \{(2l-1)v[G(A/L)] - v(AG)\} / (l-1)^2$$

Considerando que la variable respuesta sigue una distribución normal y los resultados (13) y (16) se obtendrá que cuando años se anida en localidades,

$$(ar)^2 v[\sigma_{g/2}^2] = v[GL] + v[G(A/L)] + 2\{E[G(A/L)]^2 / [(l-1)(a-1)(g-1)l]\}$$

Relacionando este resultado con el de la varianza del estimador de $v(GL)$ del modelo (3) se obtiene que

$$v[\sigma_{g/2}^2] = v[\sigma_{g/3}^2] + 2\{E[G(A/L)]^2 / [(l-1)(a-1)(g-1)a^2r^2]\}$$

La mayor varianza de σ_{g2}^2 y $\sigma_{g/2}^2$, hace que se incremente la probabilidad de tener estimaciones más alejadas del valor paramétrico sujeto a estimación, promoviendo la toma de decisiones equivocadas en la planeación y/o desarrollo de un programa de mejoramiento genético. Por ejemplo, una estimación demasiado elevada de $\sigma_{g/}$ puede inducir a la decisión de una evaluación de genotipos en una serie de localidades mayor que la requerida y viceversa.

Considerar que A y L son cruzados cuando A se anida en L también afecta la estimación de la heredabilidad, un cociente cuyo numerador es el estimador de la varianza genética y cuyo denominador es el estimador de la varianza fenotípica $[\sigma_f^2]$. Respecto a esta varianza, cuando se estima con base en el modelo (2), el estimador es de la forma

$$\sigma_{f2}^2 = \sigma_{g2}^2 + \frac{\sigma_{ag2}^2}{a} + \frac{\sigma_{g/2}^2}{l} + \frac{\sigma_{ag/2}^2}{al} + \frac{\sigma^2}{alr}$$

De acuerdo con (4), (6), (8) y (10), esta varianza también se puede expresar como

$$\sigma_{f2}^2 = \frac{G-GL}{lar} - \frac{AG-G(A/L)}{ar(l-1)} + \frac{AG-G(A/L)}{ra(l-1)} + \frac{[GL-G(A/L)]}{ral} +$$

$$\frac{AG-G(A/L)}{(l-1)arl} + \frac{[G(A/L)-E_r]}{ral} + \frac{[G(A/L)-AG]}{(l-1)arl} + \frac{E_r}{alr}$$

Pero esta expresión se reduce a

$$\sigma_{f2}^2 = \frac{G-GL}{lar} + \frac{GL-G(A/L)}{ral} + \frac{G(A/L)-E_r}{ral} + \frac{E_r}{alr}$$

$$= \sigma_{g3}^2 + \frac{\sigma_{g/3}^2}{l} + \frac{\sigma_{g(a/l)}^2}{al} + \frac{\sigma^2}{alr}$$

$$= \sigma_{f3}^2$$

en donde σ_{f3}^2 es el estimador de la varianza fenotípica del modelo (3).

Este resultado no fue sorprendente pues si bien los modelos (2) y (3) explican en forma diferente los resultados del experimento, los datos y, en consecuencia, las medias fenotípicas no dependen del modelo. Éstas son las mismas en un modelo y en otro. La varianza del estimador de σ_g^2 en cambio, es superior cuando los años se anidan en las localidades pero la estimación se realiza considerando que la relación es de cruzamiento. Ésta repercutirá en una mayor fluctuación en las estimaciones de la heredabilidad, pudiéndose, en consecuencia, generar falsas expectativas en relación al éxito que se obtendría en un programa de selección.

CONCLUSIONES

La consideración de que años (A) está cruzado con localidades (L) [Modelo (2)] cuando en realidad hay anidamiento de A en L produce

efectos indeseables en los estimadores de la varianza de los efectos de los genotipos (G) y de sus interacciones (GL, AG y AGL): Si bien los estimadores de las varianzas de G y GL resultan ser insesgados, tienen mayor varianza que los generados en el modelo en que sí se considera que A está anidado en L [Modelo (3)]. Por otra parte, los estimadores de las varianzas de las interacciones AG y AGL tienen un valor esperado igual a cero y a la varianza de la interacción entre genotipos y años anidados en localidades, respectivamente.

La suposición de que A y L están cruzados cuando en realidad A está anidado en B produce estimadores de las varianzas de G y de GL cuyas varianzas son mayores que los de los estimadores que se genera con base en el modelo en que sí se considera que A anida a L. Esta mayor varianza debe incrementar la probabilidad de tomar decisiones equivocadas respecto a la planeación y/o el desarrollo de un programa de mejoramiento genético.

BIBLIOGRAFÍA

- Addelman, S. 1970.** Variability of treatments and experimental units in the design and analysis of experiments. *Journal of the American Statistical Association* 65:1095-1108.
- Anderson, V.L. and R.A. McLean. 1974.** Design of Experiments. A Realistic Approach. Marcel Dekker, New York. 418 p.
- Lindman, H.R. 1992.** Analysis of Variance in Experimental Designs. Springer Verlag. 531 p.
- Sahagún C., J. 1993.** Funcionalidad de cuatro modelos para las evaluaciones genotípicas en series de experimentos. *Revista Fitotecnia Mexicana* 16:161-171.
- Sahagún C., J. 1994.** Evaluación de genotipos en series de experimentos: Diferencia en parámetros genéticos generados en dos modelos. *Revista Fitotecnia Mexicana* 17:116-125.
- Satterthwaite, F.E. 1946.** An approximate distribution of variance components. *Biometrics Bulletin*, 2, pp. 110-114.
- Searle, S.R. 1971.** Linear Models. Wiley, New York. 532 p.
- Searle, S.R. 1987.** Linear Models for Unbalanced Data. Wiley, New York. 536 p.
- Snedecor, G.W. and W.G. Cochran. 1980.** Statistical Methods. Seventh edition. Iowa State University Press. Ames, Ia. 507 p.
- Wilk, M.B. and O. Kempthorne. 1955.** Fixed, mixed and random models in the analysis of variance. *Journal of the American Statistical Association* 50:1144-1167.