

EVALUACION DE GENOTIPOS EN SERIES DE EXPERIMENTOS: DIFERENCIAS EN PARAMETROS GENETICOS GENERADOS EN DOS MODELOS

EVALUATION OF GENOTYPES IN SERIES OF EXPERIMENTS: DIFFERENCES AND GENETIC PARAMETERS GENERATED FROM TWO MODELS

Jaime Sahagún Castellanos¹

RESUMEN

Debido a la importancia de la inevitable fase de evaluación de genotipos cuando la meta es el desarrollo de cultivares superiores, el fitomejorador debe estar consciente del efecto que, en términos de parámetros genéticos, tiene en la calidad del análisis la correspondencia entre realidad y modelo que la representa. En este estudio se presenta un análisis preliminar sobre los efectos que en estimación de componentes de varianza, pruebas de hipótesis, heredabilidad, etc. surgen debido a la utilización de un modelo no apropiado. Los resultados indican que cuando se utiliza un modelo en el que se considera que el factor años (A) está anidado en "localidades" (L), cuando en realidad A y L son factores cruzados se producen estimadores sesgados para los componentes de varianza de genotipos (G) y de la interacción GL, así como para la heredabilidad. Además, en las pruebas de hipótesis de GL y G se produce un valor de F mayor y menor, respectivamente, que lo que sería bajo el modelo apropiado.

PALABRAS CLAVE ADICIONALES

Componentes de varianza, heredabilidad, pruebas de hipótesis, series de experimentos, modelo lineal.

SUMMARY

Due to the importance of the unavoidable genotypic evaluation stage when the development of superior cultivars is the goal, plant breeders should be aware of the effects that, in terms of genetic parameters, has in the quality of the analysis the

correspondence between reality and model representing it. In this study a preliminary analysis on the effects that show up in the estimation of variance components, hypothesis testing, heritability, etc. due to the use of a wrong model is made. Results indicate that when a model in which years (Y) is nested by locations (L) is used when Y and L are actually crossed factors, biased estimators for the variance components associated to genotypes (G), the GL interaction and heritability are produced. Furthermore, when testing the hypothesis related to GL and G the F values were higher and lower, respectively, relative to what would be under the appropriate model.

ADDITIONAL INDEX WORDS

Variance components, heritability, hypothesis testing, series of experiments, linear model.

INTRODUCCION

La evaluación de los genotipos a través del tiempo y del espacio es una condición indispensable para estimar objetivamente su auténtico potencial agronómico y de rendimiento. Sin embargo, el análisis de varianza de información producida en series de experimentos de evaluación genotípica se realiza de maneras muy diferentes, con resultados también diferentes. Mientras que muchos autores como Nguyen *et al.* (1980) y Grey (1982) consideran que los factores "años" y "localidades" son cruzados y que el factor "bloques" se encuentra anidado en "localidades", otros (Campbell y Lafever, 1976; McIntosh, 1983) consideran que si bien

¹ Departamento de Fitotecnia, UACH; CP 56230, Chapingo, Méx.

"años" y "localidades" son cruzados, bloques debe estar anidado en las combinaciones de niveles de años y localidades. Más recientemente, Sahagún (1990) en un estudio específico sobre el tópico afirma que, como los agentes del ambiente asociados con un mismo año típicamente inciden de manera diferencial en las localidades, la consideración de que años y localidades son cruzados es errónea y sugiere el uso de un modelo con "años" anidado en "localidades" para situaciones típicas de agricultura de temporal con cultivos anuales.

En un estudio posterior, Sahagún (1993) examina las diferencias entre los tres modelos anteriores, pero esto lo hace más bien en términos de las diferencias en las formas de realizar las pruebas de hipótesis y las características importantes de éstas, ello sin hacer referencia al efecto diferencial de los análisis y su incidencia en el éxito del mejoramiento genético de los cultivos, aspecto éste de mayor significado para el fitomejorador.

Dado que el modelo que considera que los factores "años" y "localidades" son cruzados es el más frecuentemente utilizado y que el valor de la información obtenida del análisis depende importantemente del grado de correspondencia entre situación real y modelo, este estudio fue concebido con el fin fundamental de estudiar el efecto de analizar la información utilizando como base el modelo en que se considera que "años" está anidado en "localidades" cuando en realidad se debiera utilizar el modelo en que "años" y "localidades" son cruzados. Dicho efecto se medirá en términos de diferencias en estimaciones de componentes varianza y pruebas de hipótesis que involucran al factor "genotipos", así como las implicaciones de estas diferencias en otros indicadores genéticos de gran interés para el fitomejorador, como lo es la heredabilidad, por ejemplo.

METODOS

Gran parte de la metodología a utilizar en este estudio ya ha sido descrita en general en reportes previos (Sahagún, 1990). A continuación se presenta una síntesis de ella.

Como en las series típicas de experimentos de campo, en este estudio se considerará la realización de un experimento en cada combinación de a años y l localidades. Cada experimento será conducido en un diseño en bloques al azar con r repeticiones y será utilizado para evaluar g genotipos, los mismos en todos los experimentos.

Un modelo que frecuentemente se utiliza para representar la situación experimental anterior es el que considera que los factores "años" (A) y "localidades" (L) son factores cruzados, que el factor "bloques" (B) está anidado tanto en A como en L y que el factor genotipos (G) está cruzado tanto con L como con A. Con estas consideraciones, Y_{ijkm} , el valor correspondiente al genotipo i en la repetición j de la localidad k durante el año m , puede ser explicado, omitiendo detalles obvios, en la forma

$$Y_{ijkm} = \mu + G_i + (B/LA)_{jkm} + L_k + (GL)_{ik} + A_m + (GA)_{im} + (LA)_{km} + (GLA)_{ikm} + E_{ijkm}. \quad (1)$$

El segundo modelo a considerar es uno en el que se reconoce que los factores ambientales asociados con los años (cantidad y distribución de precipitación, patrón de temperaturas, vientos, granizadas, etc.) pueden ocurrir de manera diferente de una localidad a otra dentro de un mismo año cronológico, en tanto que los componentes ambientales básicos asociados con las localidades se mantienen esencialmente iguales de un año a otro. Este modelo es de la forma

$$Y_{ijkm} = \mu + G_i + (B/A/L)_{ijkm} + (A/L)_{km} + (G(A/L))_{ikm} + L_m + (GL)_{im} + E_{ijkm} \quad (2)$$

Para completar la definición de los modelos es necesario determinar la naturaleza de los factores. En este trabajo se considerará que los factores A, L y B son de efectos aleatorios en tanto que con respecto al factor G se presentará un análisis considerándolo como factor de efectos fijos y otro análisis considerándolo como factor de efectos aleatorios. Además, el efecto del nivel p-ésimo, X_p , de cualquier factor o interacción aleatoria (X) de n niveles será considerado como una variable aleatoria normal e independientemente distribuida con media cero y varianza $V(x)$ ($p=1,2,\dots, n$).

Con relación a la estimación de componentes de varianza, se considerará la aplicación del método de momentos (también conocido como de análisis de varianzas). Este consiste en establecer un sistema de ecuaciones entre los cuadrados medios y sus valores esperados correspondientes. La solución a este sistema, en donde las incógnitas son los componentes de varianza en cuestión, produce los estimadores correspondientes.

Las esperanzas de cuadrados medios serán determinadas según las reglas que describen autores como Bennett y Franklin (1954), Snedecor y Cochran (1982) y Anderson y McLean (1974).

Con relación a la prueba de hipótesis de un efecto o interacción se utilizará el procedimiento usual del cociente de dos cuadrados medios en donde el numerador es el cuadrado medio correspondiente al efecto o interacción bajo prueba y el denominador un cuadrado medio que difiera del numerador sólo por que aquél no contiene el componente de varianza, o función cuadrática de

efectos (pseudovarianza) correspondiente a la fuente de variación en prueba. Cuando esto no sea posible se recurrirá al uso de una prueba aproximada según el procedimiento establecido por Satterthwaite (1946).

Ahora bien, la heredabilidad a considerar será el cociente cuyo numerador es el estimador del componente de varianza correspondiente a los genotipos y cuyo denominador es el estimador de la varianza entre las medias fenotípicas.

RESULTADOS Y DISCUSION

Para el modelo (1), en el que los factores "años" (A), "localidades" (L) y "genotipos" (G) se consideran cruzados y el factor "bloques" (B) anidado en A y L, se genera un análisis de varianza cuyos aspectos más importantes, desde el punto de vista de los objetivos de este trabajo, se presentan en el Cuadro 1. En este cuadro se muestra la importancia de considerar a G como factor de efectos fijos o de efectos aleatorios; para el primer caso todas las pruebas de hipótesis en que no está involucrado el factor G, son pruebas exactas. En cambio, cuando G es factor de efectos aleatorios todas estas pruebas, salvo la de bloques, resultan ser aproximadas. Con respecto a las fuentes de variación en que está involucrado el factor G no se presenta ninguna diferencia entre los resultados obtenidos con un modelo y otro en términos de las formas de probar hipótesis.

En el modelo (2) se considera que el factor B se encuentra anidado en el factor A y éste en el L y que G está cruzado con L. Las esperanzas de cuadrados medios y pruebas de hipótesis que se generan en este modelo se muestran en el Cuadro 2. En el caso del modelo mixto (cuando g es el único factor de efectos fijos) todas las pruebas de hipótesis resultan ser exactas. Sin embargo,

Cuadro 1. Esperanzas de cuadrados medios y pruebas de hipótesis para el modelo (1).

FV	Esperanzas de cuadrados medios ¹										Pruebas de F	
	CM	E	GAL	GL	GA	G	(B/AL)	AL	L	A	A ²	M ³
Años (A)	C ₁	p	p ⁴		p ⁴		p	p		p	(C ₁ +C ₈)/(C ₃ +C ₆)	C ₁ /C ₃
Localidades (L)	C ₂	p	p ⁴	p ⁴			p	p	p		(C ₂ +C ₈)/(C ₃ +C ₇)	C ₂ /C ₃
AL	C ₃	p	p ⁴				p	p			(C ₃ +C ₉)/(C ₄ +C ₈)	C ₃ /C ₄
Bloques/AL	C ₄	p				p	p				C ₄ /C ₉	C ₄ /C ₉
Genotipos	C ₅	p	p	p	p						(C ₅ +C ₈)/(C ₆ +C ₇)	(C ₅ +C ₈)/(C ₆ +C ₇)
GA	C ₆	p	p		p						C ₆ /C ₈	C ₆ /C ₈
GL	C ₇	P	P	P							C ₇ /C ₈	C ₇ /C ₈
GAL	C ₈	P	P								C ₈ /C ₉	C ₈ /C ₉
Error (E)	C ₉	P										

¹ varianzas o pseudovarianzas de los factores indicados² modelo aleatorio³ modelo mixto (sólo G es fijo)⁴ p = presencia del término; p⁴ = presencia sólo en el modelo aleatorio.

Cuadro 2. Esperanzas de cuadrados medios y pruebas de hipótesis del modelo (2).

FV	Esperanzas de cuadrados medios ¹						Pruebas de F			
	CM	E	G(A/L)	GL	G	B/A/L	A/L	L	A ²	M ³
Localidades (L)	C ₁	P ⁴	P ⁴	P ⁴		P	P	P	(C ₁ +C ₆)/(C ₂ +C ₃)	C ₁ /C ₃
Años (A)/L	C ₂	P ⁴	P ⁴			P	P	P	(C ₂ +C ₇)/(C ₃ +C ₆)	C ₂ /C ₃
Bloques (B)/A/L	C ₃					P			C ₃ /C ₇	C ₃ /C ₇
Genotipos (G)	C ₄	P	P	P	P				C ₄ /C ₅	C ₄ /C ₅
GL	C ₅	P	P	P					C ₅ /C ₆	C ₅ /C ₆
G(A/L)	C ₆	P	P						C ₆ /C ₇	C ₆ /C ₇
Error (E)	C ₇	P								

¹ varianzas o pseudovarianzas de los factores indicados² modelo aleatorio³ modelo mixto⁴ p = presencia del término; p⁴ = presencia sólo en el modelo aleatorio.

cuando el modelo es aleatorio, las pruebas de hipótesis correspondientes a los factores A y A/L son sólo aproximadas. Además, con base en el método del análisis de varianza, resulta que los estimadores de los componentes de varianza asociados a los factores GL y G(A/L) son exactamente los mismos, independientemente de si el modelo es aleatorio o mixto. El análisis de varianza del modelo (2) y el del modelo (1) ya fueron reportados en otro estudio (Sahagún, 1993). Su presentación en esta ocasión obedece a la necesidad de su consideración para un análisis más específico que se muestra adelante. Tanto el modelo (1) como el (2) podrían ser utilizados para series de experimentos en que se evalúan genotipos correspondientes a cultivos anuales. En el contexto de estos dos modelos, a continuación se hará referencia al tópico de estimación de los componentes de varianza de interés en el ámbito del problema de la evaluación genotípica considerando, obviamente, que los modelos (1) y (2) son modelos aleatorios. En el Cuadro 3 se muestran los estimadores de los componentes de varianza que corresponden al factor G y a las interacciones que involucran a este factor.

De acuerdo con la información del Cuadro 3 no existe coincidencia en ningún sentido entre la estimación de componentes de varianza que se generan en los modelos (1) y (2). Así, mientras que en el modelo (2), por no incluir los términos GAL y GL, el análisis de varianza no permite la estimación de los componentes de varianza asociados a GAL ($V(GAL)$) y a GL ($V(GL)$), ésta sí es posible en el modelo (1). Por otra parte, para los componentes de varianza presentes en ambos modelos, los asociados al factor G y a la interacción entre los factores G y L, las estimaciones correspondientes, dado un conjunto de datos, no necesariamente coincidirán. Un análisis de las relaciones de magnitud para estos dos casos resulta interesante. Mientras que para el modelo (1) el estimador del componente de varianza asociado a la interacción GL es de la forma $[CM(GL) - CM(GAL)]/ar$, en el modelo (2) el estimador de este componente tiene adicionalmente, con signo negativo, al término $[CM(GA) - CM(GAL)]/lar$. Esto es así porque, de acuerdo con las expresiones para sumas de cuadrados (SC),

Cuadro 3. Estimadores de componentes de varianza según los modelos (aleatorios) (1) y (2).

Factor ó interacción	M o d e l o	
	1	2
GAL	$[CM(GAL) - CM(E)]/r$	—
GL	$[CM(GL) - CM(GAL)]/ar$	$[CM(GL) - CM(G(A/L))]/ar$
GA	$[CM(GA) - CM(GAL)]/lr$	—
G(A/L)	—	$[CM(G(A/L)) - CM(E)]/r$
G	$[CM(G) - CM(GL) - CM(GA) + CM(GAL)]/lar$	$[CM(G) - CM(GL)]/lar$

[†] CM = cuadrado medio

$$\begin{aligned}
 CM [G(A/L)] &= \frac{SC [G(A/L)]}{(g-1)(a-1)l} \\
 &= \frac{SC(GAL) + SC(GA)}{(g-1)(a-1)l} \\
 &= \frac{SC(GAL)(l-1)}{(g-1)(a-1)(l-1)l} + \frac{SC(GA)}{(g-1)(a-1)l} \\
 &= \frac{(l-1)CM(GAL)}{l} + \frac{CM(GA)}{l} \\
 &= \frac{CM(GAL) - CM(GAL)/l}{1} + \frac{CM(GA)}{l}
 \end{aligned}$$

Para el caso de la componente de varianza asociada al término G la situación es un tanto a la inversa en el sentido de que mientras que el estimador de esta componente es $[CM(G) - CM(GL)]/lar$ en el modelo (2), en el modelo (1) el estimador tiene adicionalmente, con signo negativo, a la componente $[CM(GA) - CM(GAL)]/lar$. Cuando los factores A y L son cruzados, dado que los componentes de varianzas son no negativos, la relación de sus esperanzas (E) de cuadrados medios es

$$E [CM(GA)] \geq E [CM(GAL)],$$

lo cual implica que, en términos generales, se obtengan mayores estimaciones de V(G) en el modelo (2) que las obtenidas en el modelo (1). Esto es, si el modelo (1) fuera el adecuado, con el uso del modelo (2) se tendería a sobreestimar a la componente de varianza asociado a G.

Ahora bien, por lo que respecta al componente de varianza asociado a la interacción entre genotipos y localidades (GL), la diferencia entre los estimadores producidos en ambos modelos también depende de la diferencia entre CM(GA) y CM(GAL). Un

sencillo análisis indicará que cuando el modelo (1) es apropiado, el uso del modelo (2) tendería a producir subestimaciones de esta componente.

Concretando, si los factores A y L fueran realmente cruzados, el modelo (1) produciría estimadores insesgados para V(G) y V(GL) en tanto que los estimadores que se generarían usando el modelo (2) serían sesgados con sesgos de $-V(GA)/a$ y $V(GA)/a$, respectivamente. Formalmente, la justificación de esta aseveración se basa en el resultado que se muestra a continuación:

$$\begin{aligned}
 E [(CM(GAL) - CM(GA))/lar] \\
 &= [V(E) + r V(GAL) - V(E) - r V(GAL) - \\
 &\quad r l V(GA)]/lar \\
 &= -V(GA)/a
 \end{aligned}$$

Por otra parte, los estimadores para V(G) y V(GL) generados con el modelo (2), además de ser sesgados tendrían menor y mayor varianza, respectivamente, que las auténticas (las correspondientes a los estimadores producidos en el modelo (1).

El impacto de las diferencias entre los dos modelos también se hace sentir en las pruebas de hipótesis. En adición a las consideraciones ya hechas sobre este tópico resulta interesante el análisis de las pruebas de hipótesis asociadas a fuentes de variación comunes en ambos modelos. La interacción GL requiere de una prueba diferente en cada caso (Cuadro 4). En tanto que en el modelo (1) el cociente utilizado es $CM(GL)/CM(GAL)$, en el modelo (2) se utiliza el cociente $CM(GL) / CM(G(A/L))$. Dado que, como ya se demostró,

$$CM(G(A/L)) = CM(GAL) + [CM(GA) - CM(GAL)]/l,$$

Cuadro 4. Pruebas de hipótesis asociadas a G y a GL

Compo- nente	Modelo	
	1	2
GL	$\frac{CM(GL)}{CM(GAL)}$	$\frac{CM(G)}{CM(G(A/L))}$
G	$\frac{CM(G) + CM(GAL)}{CM(GA) + CM(GL)}$	$\frac{CM(G)}{CM(GL)}$

entonces para un conjunto dado de datos no necesariamente se tiene que producir un mismo valor para ambos cocientes, los denominadores pueden diferir. Si el modelo (1) fuera el apropiado, como ya se ha expresado,

$$E [CM(GA)] \geq E [CM(GAL)],$$

y el cociente asociado a la prueba de la misma hipótesis que se produciría utilizando el modelo (2) tendería a ser menor que la correspondiente a la misma prueba en el modelo (1). Cabe señalar, sin embargo, que la distribución F involucrada tendría un número mayor de grados de libertad en el denominador para el caso del modelo (2).

Ahora bien, con relación a la prueba de hipótesis asociada al factor G (genotipos), dada la estructura de los cocientes, resulta que para el caso del modelo (1) debería esperarse una tendencia a producir valores de F menores que los que se producirían si se usara el modelo (2) cuando el modelo (1) fuera el indicado. Los grados de libertad de las F's, debe considerarse, serían diferentes en ambos casos también. En la realidad estas diferencias pueden tener implicaciones.

Tanto la prueba de hipótesis para G como la prueba de hipótesis para GL son de gran

significado en el ámbito del mejoramiento genético. Por ejemplo, cuando la interacción GL resulta ser significativa eso podría representar para el fitomejorador la necesidad de seguir realizando la evaluación de genotipos en varias localidades. Un componente de varianza $V(GL)$ relativamente pequeño, por el contrario, sería un fuerte argumento para intensificar la evaluación genotípica a través del tiempo más que del espacio.

En otro aspecto de gran importancia para el fitomejorador, la heredabilidad, los modelos (1) y (2) también ejercen efectos diferenciales. En particular, con relación a la varianza entre las medias fenotípicas, las estimaciones producidas en los modelos (1) y (2), y otros similares, dado un conjunto de datos cualquiera, siempre serán idénticas. El valor de esta estimación siempre será el correspondiente al cuadrado medio debido a G (genotipos) dividido por el producto "lar" (Cuadro 5). En contraste, el estimador del componente de varianza asociado a los genotipos es diferente de un modelo a otro. Si el modelo (1) fuera el indicado, y si $\hat{V}(G_1)$ y $\hat{V}(G_2)$ representaran a los estimadores de la varianza debida al factor "genotipos" en los modelos (1) y (2), respectivamente, entonces, de acuerdo con la discusión de los resultados del Cuadro 3, $V(G_1) \leq V(G_2)$ y, por esta razón, las estimaciones de heredabilidad obtenidas con el modelo (2) tenderían a ser mayores que las obtenidas con el modelo (1). Esto es, si el modelo (1) fuera el correcto, el uso del modelo (2) produciría un sesgo en la estimación de la heredabilidad.

De acuerdo con los resultados anteriores, la respuesta a la selección, la cual es directamente proporcional a la heredabilidad, presentará cambios de un modelo a otro de la misma naturaleza que los descritos para la heredabilidad.

Cuadro 5. Varianzas entre las medias fenotípicas según los modelos (1) y (2).

Modelo	V a r i a n z a								
(1)	V(G)	+	$\frac{V(GA)}{a}$	+	$\frac{V(GL)}{l}$	+	$\frac{V(GAL)}{al}$	+	$\frac{V(E)}{alr}$
(2)	V(G)	+	$\frac{V(G(A/L))}{al}$	+	$\frac{V(GL)}{l}$	+	$\frac{V(E)}{alr}$		

La heredabilidad, u otros conceptos estrechamente relacionados con ella como la respuesta a la selección, son parámetros que por si mismos orientan al fitomejorador para la toma de decisiones muy importantes en el proceso de desarrollo de cultivares superiores. Por ejemplo, el uso de estimaciones de respuesta a la selección y de la heredabilidad puede orientar efectivamente para decidir sobre el método de mejoramiento a utilizar. Evidentemente, la calidad de la orientación dependerá importantemente de lo acertada que haya sido la representación, mediante un modelo estadístico, de la situación experimental.

De las consideraciones anteriores se han derivado algunos resultados considerando que la situación experimental es acorde con el modelo (1). Esta consideración, que pudiendo no ser realista en muchos casos, permite advertir en forma específica, con criterios genotécnicos, la importancia que para el fitomejorador tiene la selección de un modelo u otro. Si, en cambio, el modelo (2) fuera acorde con la situación real en turno y el (1) fuera el que se utilizara, también habría problemas. Sin ser el objetivo de este estudio el análisis correspondiente a esta situación, sólo con fines de evidenciar las dificultades se hará referencia a un problema de interpretación. Por ejemplo, dado que en el modelo (1) se incluyen los términos de interacción GA y GAL, si

en el análisis se detectara significancia, digamos en GA, sería razonable pensar que ésta pudo haber sido alcanzada por la variación generada por factores meteorológicos (cantidad y distribución de precipitación, ocurrencia de heladas, granizadas, etc.) que se presenta dentro de cada año cronológico a través de las diferentes localidades, en tanto que en la interpretación se lleva el entendimiento tácito de que en cada año cronológico los factores meteorológicos inciden igualmente en todas las localidades.

BIBLIOGRAFIA

- Anderson, V.L. and R.A. McLean. 1974. Design of Experiments. A Realistic Approach. Marcel Dekker, New York. 418 pp.
- Bennett, C.A. and N.L. Franklin. 1954. Statistical Analysis in Chemistry and the Chemical Industry. Wiley, New York.
- Campbell, L.G. and H.N. Lafever. 1976. Genotype x epvIRONMENTAL interactions in soft red winter wheat yield test. Crop Sci. 17:604-608.
- Grey, E. 1982. Genotype x environment interactions and stability analysis for forage yield of orchardgrass clones. Crop Sci. 22:19-23.
- McIntosh, M.S. 1983. Analysis of combined experiments. Agron. J. 75: 153-155.

Nguyen, H.T., D.A. Sleper, and K.L. Hunt. 1980. Genotype x environment interaction and stability analysis for herbage yield of tall fescue synthetics. Crop Sci. 20:221-224.

Sahagún C., J. 1990. La utilidad del análisis de varianza en el estudio de la interacción entre genotipos y ambientes. Xilonen 1:21-32.

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

Snedecor, G.W. and W.G., Cochran. 1980. Statistical Methods. Seventh edition Iowa State University Press. Ames Ia. 507 pp.

_____. 1993. Funcionalidad de cuatro modelos para las evaluaciones genotípicas en series de experimentos. Fitotecnía Mexicana 16:161-171.