

MEJOR PREDICTOR LINEAL INESGADO FAMILIAR EN EXPERIMENTOS PARCIALES DE CRUZAS DIALÉLICAS SIN EFECTOS MATERNOS

BEST LINEAR UNBIASED FAMILIAR PREDICTOR FOR PARTIAL DIALLEL EXPERIMENTS WITHOUT MATERNAL EFFECTS

Guadalupe Hernández Lira¹, Ángel Martínez Garza^{1*}, Ángel Agustín Mastache Lagunas² y Gilberto Rendón Sánchez¹

¹ Programa en Estadística, Instituto de Socioeconomía, Estadística e Informática, Colegio de Postgraduados, Montecillo, Texcoco, Edo. De México. Tel. y Fax (01-595)10194. Correo electrónico: angel@colpos.mx ² Centro de Estudios Profesionales, Colegio Superior Agropecuario del Edo. de Guerrero, Iguala, Gro.

* Autor responsable

RESUMEN

Cuando en un experimento de cruas dialélicas se ensaya un subconjunto de todas las cruas simples que pueden generarse a partir de un conjunto básico de p progenitores, se obtiene la clase de los experimentos dialélicos parciales. De éstos, los más útiles ocurren cuando cada progenitor se involucra en un número de cruas $s < p$. Si, adicionalmente, los progenitores se agrupan en dos o más subconjuntos o familias, donde los miembros de un subconjunto particular se distinguen por ciertas características comunes, surge el problema de estimar los efectos de familia o de grupo. Este problema se resuelve en esta contribución, mediante los mejores predictores lineales e inesgados empíricos de los efectos familiares, en ausencia de efectos maternos. Como complemento a la investigación, se construye un algoritmo computacional en SAS/IML para efectuar el análisis de estos experimentos, que produce también los correspondientes estimadores de los efectos familiares.

Palabras clave: Diseños dialélicos parciales, mejores predictores lineales, mejores predictores inesgados empíricos, aptitud combinatoria general, efecto familiar, algoritmo computacional.

SUMMARY

When an experiment of diallel crosses involves a subset of all possible simple crosses, which can be generated from a basic set of p parents, a partial diallel experiment is obtained. The most useful of these experiments are those in which every parent is involved in a number $s < p$ of crosses. If, in addition, the parents are grouped into two or more subsets or families, where the members of a particular subset are distinguished for some common characteristics, then the problem arises when estimating the effects of family or group. This problem is solved in this paper, by means of an empirical best linear unbiased predictor of the family effects, in absence of maternal effects. As a complement of this research, a computational algorithm in SAS/IML commands is also obtained, in order to analyse the experiments, giving at the same time the corresponding estimators of the family effects.

Index words: Partial diallel designs, best linear predictors, empirical best unbiased predictors, general combining abilities, familiar effects, computational algorithm.

INTRODUCCIÓN

En los experimentos de cruas dialélicas, los esquemas completos de Griffing son útiles cuando el número de líneas progenitoras es pequeño; sin embargo, cuando este número es elevado es difícil preparar, establecer y conducir los trabajos de campo. Una alternativa consiste en emplear los diseños parciales de cruas dialélicas. Estos esquemas ensayan un subconjunto del conjunto total de cruas que es posible formar entre los progenitores básicos y permiten la inclusión de un mayor número de líneas progenitoras, y balancear los niveles de precisión en la estimación de los componentes de la variación.

La mayor ventaja de los experimentos dialélicos parciales es proporcionar una herramienta más elástica al genetista, quien en sus programas de mejoramiento de plantas y animales puede tener un número grande de líneas progenitoras. Estos experimentos pueden ser simétricos o asimétricos; los primeros, requieren que cada progenitor se involucre en el mismo número de cruas; los últimos requieren que al menos uno de los progenitores participe en un número diferente de cruas. Los más útiles son, obviamente, los que cumplen ciertas condiciones de simetría. Algunas soluciones de diseño cuando no se ensayan las cruas recíprocas, son las propuestas por Kempthorne y Curnow (1961), Fyfe y Gilbert (1963) y Rojas (1973), quienes abordaron el análisis y la estimación de parámetros en estos diseños a partir de un modelo de

efectos fijos; Martínez (1983, 1991) resume estos procedimientos.

En las situaciones en las que los efectos de los factores involucrados en el modelo se consideran aleatorios o mixtos, es posible derivar los mejores predictores lineales e insesgados (MPLI), mediante la metodología desarrollada por Henderson (1975 y 1984). Es frecuente que los investigadores formen grupos o familias de progenitores, como lo hacen Cervantes *et al.* (1999) y Ron *et al.* (1999), lo que implica considerar la estimación de los efectos de grupo o familia.

Si los componentes de la varianza involucrados son conocidos en estos predictores, el estimador de la aptitud combinatoria general dependerá de la matriz diseño asociada a los progenitores, la cual es muy variable. La dificultad es mayor cuando se desconocen los componentes de varianza y tienen que ser estimados para obtener los MPLI empíricos de los efectos de progenitor. Lo mismo ocurre con el estimador del efecto familiar o de grupo, cuando los progenitores se agrupan en subconjuntos, donde los miembros de un subconjunto particular se caracterizan por poseer características comunes, lo que da lugar a matrices diseño muy específicas para cada caso y no se tiene una forma matemática sencilla de generalización.

En este trabajo se derivaron los MPLI empíricos de los efectos familiares o de grupo, en experimentos dialélicos parciales sin efectos maternos, cuyos resultados se expresan en notación matricial; la investigación se complementa con un algoritmo computacional escrito en comandos SAS/IML, para aplicar la técnica de estimación.

METODOLOGÍA

El modelo lineal de efectos mixtos, puede representarse de la siguiente manera

$$y = X\beta + Zu + e \tag{Ec. 1}$$

donde y es un vector $n \times 1$ de observaciones, X y Z son matrices diseño conocidas de dimensiones $n \times f$ y $n \times p$, respectivamente, β es un vector $f \times 1$ de efectos fijos desconocidos, u es un vector $n \times 1$ no observable de efectos aleatorios, y e es un vector $n \times 1$ no observable de términos de error. Además, $u \sim (0, G\sigma_e^2)$ y $e \sim (0, R\sigma_e^2)$, con

$$Var \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \sigma_e^2 \tag{Ec. 2}$$

donde σ_e^2 es una escalar positivo generalmente desconocido; G y R son matrices de constantes y usualmente $G = (\sigma_u^2 / \sigma_e^2) I$. Henderson (1963, 1973), bajo la suposición adicional de normalidad multivariable de los vectores u y e , reporta que la densidad conjunta de u y e es

$$f(y, u) = d(y|u)c(u) \tag{Ec. 3}$$

donde $c(u)$ es la densidad marginal de u y $d(y|u)$ es la distribución condicional de y dado u . Bajo estas condiciones, Henderson (1963) derivó las ecuaciones normales del modelo mixto mediante la maximización de la densidad conjunta (Ec. 3), con respecto a β y u . Explícitamente, $f(y, u)$ tiene la siguiente estructura:

$$f(y, u) \propto e^{-\frac{1}{2\sigma_e^2} (y - X\beta - Zu)' R^{-1} (y - X\beta - Zu)} \left\{ e^{-\frac{1}{2\sigma_e^2} u' G^{-1} u} \right\} \tag{Ec. 4}$$

La Ec. 4 se reduce a la forma:

$$f(y, u) \propto e^{-\frac{1}{2\sigma_e^2} [(y - X\beta - Zu)' R^{-1} (y - X\beta - Zu) + u' G^{-1} u]}$$

Al tomar logaritmos y diferenciar con respecto a β y u , se tiene sucesivamente:

$$\begin{aligned} \ln f(y, u) &= -\frac{1}{2\sigma_e^2} [(y - X\beta - Zu)' R^{-1} (y - X\beta - Zu) + u' G^{-1} u] \\ &= -\frac{1}{2\sigma_e^2} (y'R^{-1}y - y'R^{-1}X\beta - y'R^{-1}Zu - \beta'X'R^{-1}y + \beta'X'R^{-1}X\beta \\ &\quad + \beta'X'R^{-1}Zu - u'Z'R^{-1}y + u'Z'R^{-1}X\beta + u'Z'R^{-1}Zu + u'G^{-1}u) \end{aligned}$$

Por ser escalares $\beta'X'R^{-1}y = y'R^{-1}X\beta$ y $u'Z'R^{-1}y = y'R^{-1}Zu$, de aquí

$$\ln f(y,u) = -\frac{1}{2\sigma_e^2}(y'R^{-1}y - 2\beta'X'R^{-1}y + \beta'X'R^{-1}X\beta + 2\beta'X'R^{-1}Zu - 2u'Z'R^{-1}y + u'Z'R^{-1}Zu + u'G^{-1}u),$$

de donde

$$\frac{\partial \ln f(y,u)}{\partial \beta'} = -\frac{1}{2\sigma_e^2}(2X'R^{-1}X\beta + 2X'R^{-1}Zu - 2X'R^{-1}y), y$$

$$\frac{\partial \ln f(y,u)}{\partial u'} = -\frac{1}{2\sigma_e^2}(2Z'R^{-1}X\beta + 2Z'R^{-1}Zu - 2Z'R^{-1}y + 2G^{-1}u)$$

Al igualar a cero las derivadas anteriores y reordenar términos, se obtienen las ecuaciones normales para el modelo mixto:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix} \quad (\text{Ec. 5})$$

Harville y Carriquiry (1992) mencionan que en la mayoría de las aplicaciones los componentes de la varianza en *G* y *R* son desconocidos y necesitan ser estimados a partir de información experimental. El procedimiento de estimación consta de dos etapas; en la primera se estiman los componentes de la varianza y en la segunda se utilizan estos estimadores y se incluyen en las ecuaciones normales del modelo mixto (Ec. 5) y se procede a obtener el MPLI, ahora llamado MPLI empírico. En los experimentos parciales se estudian *p* progenitores, via *t* cruza directas, en que $t < \frac{p(p+1)}{2}$, o $t < \frac{p(p-1)}{2}$, si se considera o no a las autofecundaciones, respectivamente.

Se parte del supuesto de que las cruza participantes se evalúan en un diseño de bloques completos al azar. Sin pérdida de generalidad, si sólo se consideraran las medias de cruza y dado que los efectos de bloques son ortogonales con los efectos de cruza, el modelo puede escribirse como:

$$n_{ij}\bar{y}_{ij} = n_{ij}(\mu + g_i + g_j + s_{ij} + \bar{e}_{ij}), \quad 1 \leq i, j \leq p \quad (\text{Ec. 6})$$

donde $n_{ij} = 1$, si la cruza (*i, j*) participa en el experimento, y $n_{ij} = 0$, en caso contrario; \bar{y}_{ij} es la media aritmética de la característica en estudio correspondiente a la cruza (*i, j*); μ es un efecto común a todas las observaciones; *g*

es el efecto de aptitud combinatoria general del progenitor *i*; *s_{ij}* es el efecto de la aptitud combinatoria específica de la cruza (*i, j*); y \bar{e}_{ij} es el promedio sobre todos los bloques de los elementos aleatorios de error correspondientes a la cruza (*i, j*). Los términos *g_i*, *s_{ij}* y \bar{e}_{ij} se consideran como variables aleatorias normales no correlacionadas entre y dentro de ellas, con medias cero y varianzas σ_g^2, σ_s^2 y σ_e^2 respectivamente.

En general, el modelo (Ec. 6) también puede expresarse matricialmente de la siguiente forma:

$$y = j\mu + Z_p g + s + \bar{e} \quad (\text{Ec. 7})$$

donde **y** es el vector de medias de las cruza participantes en el estudio, *j* es un vector de unos, *s* y \bar{e} son los vectores de aptitud combinatoria específica y los errores, respectivamente; todos ellos, de dimensión *t* × *l*; *g* es el vector de aptitud combinatoria general *s* de dimensión *p* × *l*; *Z_p* es la matriz diseño asociada a la participación de los progenitores en las diferentes cruza, cuya dimensión es *t* × *p* y es de la forma siguiente:

$$Z_p = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ z_{t1} & z_{t2} & \dots & z_{tp} \end{bmatrix} \quad (\text{Ec. 8})$$

con $z_{ii} = 1$, si el progenitor *i* participa en la cruza *l*, $l = 1, 2, \dots, t$, y $z_{ij} = 0$, si el progenitor *i* no participa en la cruza *l*; es decir, cada hilera de *Z_p* contiene dos unos y *p*-2 ceros. Si se consideran las autofecundaciones, en la matriz anterior (Ec. 8) se incluirán *p* hileras adicionales compuestas por un dos y *p*-1 ceros.

De acuerdo con lo anterior,

$$Var(y) = E[(y - j\mu)(y - j\mu)'] , \text{ o bien:}$$

$$Var(y) = E\left[(Z_p g + s + \bar{e})(Z_p g + s + \bar{e})' \right]$$

$$= Z_p \sigma_g^2 I_p Z_p' + \sigma_s^2 I_t + \left[\frac{\sigma_e^2}{r} \right] I_t$$

$$= Z_p \sigma_g^2 I_p Z_p' + \left[\frac{\sigma_e^2 + r \sigma_s^2}{r} \right] I_t \quad (\text{Ec. 9})$$

donde **I** es una matriz identidad de tamaño *t* y *r* es el número de bloques del experimento.

Si en la Ec. 7 se considera $e^* = s + \bar{e}$, ésta queda como

$$y = j\mu + Z_p g + e^*$$

Este modelo es un caso particular del modelo de efectos mixtos definido en la Ec. 1 en el que $n = t$, $f = 1$, y sus componentes son equivalentes a $X = j$, $\beta = \mu$,

$$Z = Z_p, \quad u = g \quad \text{y} \quad e = e^*. \quad \text{Además, con } \sigma_{e^*}^2 = \left[\frac{\sigma_e^2 + r \sigma_s^2}{r} \right],$$

la matriz de varianzas y covarianzas de la Ec. 9, se puede escribir como

$$\text{Var}(y) = \sigma_g^2 Z_p Z_p' + \sigma_{e^*}^2 I_t$$

$$= \left[\begin{array}{c} \sigma_g^2 \\ \sigma_{e^*}^2 \end{array} Z_p Z_p' + I_t \right] \sigma_{e^*}^2$$

$$= [Z_p G Z_p' + I_t] \sigma_{e^*}^2$$

donde $G_p = (\sigma_g^2 / \sigma_{e^*}^2) I_p$, I_p es una matriz identidad de tamaño *p* y **R** de la Ec. 2 es igual a **I**. Bajo estas consideraciones, las ecuaciones normales del modelo mixto (Ec. 5), se reducen a

$$j'j\hat{\mu} + j'Z_p\hat{g} = j'y$$

$$Z_p'j\hat{\mu} + [Z_p'Z_p + G_p^{-1}]\hat{g} = Z_p'y \quad (\text{Ec. 10})$$

RESULTADOS

Si los componentes de varianza σ_e^2 , σ_g^2 y σ_s^2 son conocidos, el mejor predictor lineal insesgado \hat{g} de *g*, en el problema específico que se está considerando, se obtiene al resolver el sistema de ecuaciones normales (Ec. 10). Así, con la restricción, resulta que $j'Z_p\hat{g} = (p-1)\sum_{i=1}^p \hat{g}_i = 0$; por tanto,

$$\hat{\mu} = (j'j)^{-1} j'y \quad (\text{Ec. 11})$$

$$\hat{g} = (Z_p'Z_p + G_p^{-1})^{-1} (Z_p'y - Z_p'j\hat{\mu}) \quad (\text{Ec. 12})$$

La solución en Ec. 11 es equivalente a

$$\hat{\mu} = \left(\frac{1}{t} \right) j'y = \frac{y \dots}{tr} = \bar{y} \dots$$

Si $\bar{y} \dots$ se sustituye en Ec. 12, \hat{g} es equivalente a

$$\hat{g} = (Z_p'Z_p + G_p^{-1})^{-1} (Z_p'y - Z_p'j\bar{y} \dots) \quad (\text{Ec. 13})$$

Si los componentes de la varianza involucrados en G_p son conocidos, el estimador de *g* dependerá de la matriz diseño Z_p . Por ejemplo, cuando no se ensayan las autofecundaciones, la matriz $Z_p'Z_p$ estará formada en su diagonal por el número de veces en que participen cada uno de los progenitores, pero fuera de ella, por ceros y unos. Lo anterior hace que el estimador de *g* no tenga una manera matemática sencilla de generalización, y la dificultad es mayor cuando se desconocen los componentes de varianza y tienen que ser estimados para obtener el MPLI empírico; esto implica que el estimador del efecto familiar no tenga una forma matemática sencilla de ser generalizada.

Ejemplo numérico

Con fines exclusivamente ilustrativos se analizarán los datos del Cuadro 1. Se tienen seis progenitores y 12 cruzas directas: (1,2), (1,3), (1,5), (1,6), (2,3), (2,4), (2,6), (3,4), (3,5), (4,5), (4,6) y (5,6); éstas se ensayan en un experimento de dos bloques completos al azar, que constituye un experimento dialélico parcial sin efectos maternos.

Cuadro 1. Dialéctico parcial generado con datos del Dr. Aquiles Carballo Carballo¹. Se examinan rendimientos de maíz (*Zea mays* L.) en t ha⁻¹.

Cruza	i*		j		Bloque	Cruza	i		j		Bloque
	I	II	I	II			I	II			
1	1	2	6.6	3.6	8	2	6	7.3	6.0		
2	1	3	5.5	5.4	11	3	4	8.2	6.0		
3	1	5	8.0	8.2	12	3	5	4.5	5.0		
4	1	6	4.2	7.1	15	4	5	6.4	6.5		
6	2	3	7.2	7.9	16	4	6	8.2	6.0		
7	2	4	6.3	4.9	20	5	6	4.5	5.0		

*i y j son en el programa del Apéndice; I y J, los progenitores femenino y masculino, respectivamente.

En este apartado se utiliza el programa desarrollado por el Dr. Angel A. Mastache Lagunas² para la obtención de $Z_p'Z_p + G_p^{-1}$, $Z_p'y$ y $Z_p'j\bar{y}...$, así como la ayuda del procedimiento IML de SAS, para realizar el cálculo de productos matriciales y de las inversas necesarias para analizar estos datos. Después se presentan los resultados producidos por un programa SAS, para verificar si el procedimiento realizado anteriormente es el correcto. A partir de los datos del Cuadro 1, para valores medios de las cruza:

$$Z_p = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

y de aquí

$$Z_p'Z_p + G_p^{-1} = \begin{bmatrix} 4 & 1 & 1 & 0 & 1 & 1 \\ 1 & 4 & 1 & 1 & 0 & 1 \\ 1 & 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 4 & 1 & 1 \\ 1 & 0 & 1 & 1 & 4 & 1 \\ 1 & 1 & 0 & 1 & 1 & 4 \end{bmatrix} = Z_p'Z_p, \text{ dado que}$$

$$G_p = \begin{bmatrix} \sigma_g^2 \\ \sigma_g^2 \\ \sigma_{e^*}^2 \end{bmatrix} I_p = 0$$

puesto que σ_g^2 se estima por cero, como se verá en la impresión de SAS. Así mismo:

$$Z_p'y = \begin{bmatrix} 24.35 \\ 24.30 \\ 23.80 \\ 27.75 \\ 25.00 \\ 23.10 \end{bmatrix} \text{ y } Z_p'j\bar{y}... = \begin{bmatrix} 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \end{bmatrix}$$

Consecuentemente el sistema de ecuaciones normales es

$$Z_p'j\hat{\mu} + [Z_p'Z_p + G_p^{-1}]\hat{g} = Z_p'y$$

o, en términos numéricos

$$\begin{bmatrix} 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \end{bmatrix} + \begin{bmatrix} 4 & 1 & 1 & 0 & 1 & 1 \\ 1 & 4 & 1 & 1 & 0 & 1 \\ 1 & 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 4 & 1 & 1 \\ 1 & 0 & 1 & 1 & 4 & 1 \\ 1 & 1 & 0 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_4 \\ \hat{g}_5 \\ \hat{g}_6 \end{bmatrix} = \begin{bmatrix} 24.35 \\ 24.30 \\ 23.80 \\ 27.75 \\ 25.00 \\ 23.10 \end{bmatrix} \quad (\text{Ec. 14})$$

Por consiguiente se obtiene

$$\begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_4 \\ \hat{g}_5 \\ \hat{g}_6 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 1 & 0 & 1 & 1 \\ 1 & 4 & 1 & 1 & 0 & 1 \\ 1 & 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 4 & 1 & 1 \\ 1 & 0 & 1 & 1 & 4 & 1 \\ 1 & 1 & 0 & 1 & 1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 24.35 - 24.716 \\ 24.30 - 24.716 \\ 23.80 - 24.716 \\ 27.75 - 24.716 \\ 25.00 - 24.716 \\ 23.10 - 24.716 \end{bmatrix}$$

de donde, después de realizar operaciones matriciales:

$$\begin{bmatrix} \hat{g}_1 \\ \hat{g}_2 \\ \hat{g}_3 \\ \hat{g}_4 \\ \hat{g}_5 \\ \hat{g}_6 \end{bmatrix} = \begin{bmatrix} 0.24175 \\ -0.12075 \\ -0.54575 \\ 1.09175 \\ 0.05425 \\ -0.72075 \end{bmatrix}$$

Al regresar al sistema de ecuaciones normales (Ec. 14), se supondrá que los progenitores se agrupan en 3 familias, con los progenitores 1 y 2 en la familia 1, los progenitores 3 y 4 en la familia 2, y los progenitores 5 y 6 en la familia 3, cuyos estimadores de los efectos familiares se denotan como \hat{f}_1 , \hat{f}_2 y \hat{f}_3 . De aquí las ecuaciones normales toman la forma:

¹ Profesor Investigador del Programa en Genética, Instituto de Recursos Naturales, Colegio de Postgraduados.

² Colegio Superior Agropecuario del Estado de Guerrero, Iguala, Gro.

$$\begin{bmatrix} 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \\ 24.716 \end{bmatrix} + \begin{bmatrix} 4 & 1 & 1 & 0 & 1 & 1 \\ 1 & 4 & 1 & 1 & 0 & 1 \\ 1 & 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 4 & 1 & 1 \\ 1 & 0 & 1 & 1 & 4 & 1 \\ 1 & 1 & 0 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{f}_3 \end{bmatrix} = \begin{bmatrix} 24.35 \\ 24.30 \\ 23.80 \\ 27.75 \\ 25.00 \\ 23.10 \end{bmatrix}$$

Al reordenar términos y efectuar operaciones se tiene

$$\begin{bmatrix} 4 & 1 & 1 & 0 & 1 & 1 \\ 1 & 4 & 1 & 1 & 0 & 1 \\ 1 & 1 & 4 & 1 & 1 & 0 \\ 0 & 1 & 1 & 4 & 1 & 1 \\ 1 & 0 & 1 & 1 & 4 & 1 \\ 1 & 1 & 0 & 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_2 \\ \hat{f}_3 \\ \hat{f}_3 \end{bmatrix} = \begin{bmatrix} -0.366 \\ -0.416 \\ -0.916 \\ 3.034 \\ 0.284 \\ -1.616 \end{bmatrix}$$

o

$$\begin{bmatrix} 4\hat{f}_1 + \hat{f}_1 + \hat{f}_2 + \hat{f}_3 + \hat{f}_3 \\ \hat{f}_1 + 4\hat{f}_1 + \hat{f}_2 + \hat{f}_2 + \hat{f}_3 \\ \hat{f}_1 + \hat{f}_1 + 4\hat{f}_2 + \hat{f}_2 + \hat{f}_3 \\ \hat{f}_1 + \hat{f}_2 + 4\hat{f}_2 + \hat{f}_3 + \hat{f}_3 \\ \hat{f}_1 + \hat{f}_2 + \hat{f}_2 + 4\hat{f}_3 + \hat{f}_3 \\ \hat{f}_1 + \hat{f}_1 + \hat{f}_2 + \hat{f}_3 + 4\hat{f}_3 \end{bmatrix} = \begin{bmatrix} -0.366 \\ -0.416 \\ -0.916 \\ 3.034 \\ 0.284 \\ -1.616 \end{bmatrix}$$

Este último conjunto de ecuaciones conduce al sistema de ecuaciones normales para estimar los efectos familiares:

$$\begin{aligned} 10\hat{f}_1 + 3\hat{f}_2 + 3\hat{f}_3 &= -0.782 \\ 3\hat{f}_1 + 10\hat{f}_2 + 3\hat{f}_3 &= 2.118 \\ 3\hat{f}_1 + 3\hat{f}_2 + 10\hat{f}_3 &= -1.332 \end{aligned}$$

después de sumarlas por grupos familiares y de reducir términos semejantes. Al resolver este sistema de ecuaciones para \hat{f}_1 , \hat{f}_2 y \hat{f}_3 , se obtienen los correspondientes estimadores de los efectos familiares. La solución expresada en notación matricial, es:

$$\begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \\ \hat{f}_3 \end{bmatrix} = \begin{bmatrix} 10 & 3 & 3 \\ 3 & 10 & 3 \\ 3 & 3 & 10 \end{bmatrix}^{-1} \begin{bmatrix} -0.782 \\ 2.118 \\ -1.332 \end{bmatrix} = \begin{bmatrix} -0.1119 \\ 0.3024 \\ -0.1904 \end{bmatrix}$$

Se observa en general, que los predictores de los efectos familiares no son el promedio de los predictores de los miembros de la familia, lo cual es cierto para algunos casos particulares.

Como se ve, el proceso de estimación de efectos familiares en experimentos dialélicos parciales, es complicado. El programa SAS/IML del Apéndice, lo aplica con facilidad. Para el presente ejemplo ilustrativo, los resultados de SAS se reproducen en la siguiente sección.

Resultados con el algoritmo computacional

Al operar con el programa SAS mencionado, se obtienen los resultados que siguen:

ANÁLISIS FAMILIAR DE EXPERIMENTOS SIMÉTRICOS BALANCEADOS DE CRUZAS DIALÉLICAS
DISEÑO DE BLOQUES AL AZAR

DIALÉLICO PARCIAL SIMÉTRICO-SIN CRUZAS RECÍPROCAS

N	T	R	P	VARIABLE
24	12	2	6	1

CUADRO 1. ANÁLISIS DE VARIANZA.

FV	GL	SC	CM	F	Pr > F
BLOQUES	1	1.1704167	1.1704167	1.0133810	0.3357252
CRUZAS	11	22.4445830	2.0404167	1.7666525	0.1796823
ACG	5	9.9083333	1.9816667	0.9484495	0.5129745
E.FAM.	2	1.9633333	0.9816667	0.4698375	0.6463033
E.IFAM.	3	7.9450000	2.6483333	1.2675242	0.3668521
ACE	6	12.5362500	2.0893750	1.8090420	0.1865907
ERROR	11	12.7045830	1.1549621	.	.
TOTAL	23	36.3195830	.	.	.

MEDIA	CV
6.17917	17.39218

ESTIMACIÓN DE LAS COMPONENTES DE VARIANZA

VARE	VARS	VARG
1.15496	0.46721	0.00000

MU
6.17917

CUADRO 2. ESTIMACIÓN Y PREDICCIÓN DE ACG

PROG	EMCG	MPLI	EMCG+MU	MPLI+MU
1	0.24167	0.24167	6.42083	6.42083
2	-0.12083	-0.12083	6.05833	6.05833
3	-0.54583	-0.54583	5.63333	5.63333
4	1.09167	1.09167	7.27083	7.27083
5	0.05417	0.05417	6.23333	6.23333
6	-0.72083	-0.72083	5.45833	5.45833

CUADRO 3. ESTIMACIÓN Y PREDICCIÓN FAMILIAR

FAMILIA	EMCGg	MPLIg
1	-0.11190	-0.11190
2	0.30238	0.30238
3	-0.19048	-0.19048

CUADRO 4. LA MATRIZ DE COEFICIENTES C22 DE ACG

CCCC						
0.267	-0.072	-0.072	0.041	-0.072	-0.072	
-0.072	0.267	-0.072	-0.072	0.041	-0.072	
-0.072	-0.072	0.267	-0.072	-0.072	0.041	
0.041	-0.072	-0.072	0.267	-0.072	-0.072	
-0.072	0.041	-0.072	-0.072	0.267	-0.072	
-0.072	-0.072	0.041	-0.072	-0.072	0.267	

CONCLUSIONES

Como resultados de la presente investigación, se resuelve en términos generales el problema de la estimación de efectos familiares en los experimentos parciales de cruza dialélicas y se ilustra la discusión con un ejemplo. Se observa que el MPLI empírico de los efectos familiares no es necesariamente el promedio de los MPLI de los efectos de aptitud combinatoria general de los miembros del grupo. Un programa escrito en comandos SAS/IML analiza los experimentos y produce los estimadores de los efectos familiares.

BIBLIOGRAFÍA

- Cervantes S T, F Castillo G, E Domínguez C (1999) Cambios en la media y en la varianza de caracteres cuantitativos de cebada por irradiación recurrente. *Agrociencia* 33:251-299.
- Fyfe J L, N Gilbert (1963) Partial diallel crosses. *Biometrics* 19: 278-286.
- Harville D A, A L Carriquiry (1992) Classical and Bayesian prediction as applied to an unbalanced mixed linear model. *Biometrics* 48: 987-1003.
- Henderson C R (1963) Selection index and expected genetic advance. *In*: W. D. Hanson and H. F. Robinson (eds.). *Statistical Genetics and Plant Breeding*. Nat. Acad. Sci., Nat. Res. Council, Publication 982, Washington, DC. pp: 141-163.
- Henderson C R (1973) Sire evaluation and genetics trends. *In*: Proc. Anim. Breed. Genet. Symp. in Honor of Dr. Jay L. Lush. Am. Soc. Anim. Sci. Champaign, ILL. pp:10-41.
- Henderson C R (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
- Henderson C R (1984) *Applications of Linear Models in Animal Breeding*. Guelph, Canada University of Guelph. pp:462.
- Kempthorne O, R N Curnow (1961) The partial diallel cross. *Biometrics* 17:229-250.
- Martínez G A (1983) *Diseño y análisis de los experimentos de cruza dialélicas*. Centro de Estadística y Cálculo. Colegio de Postgraduados, Chapingo, México. 252 p.
- Martínez G A (1991) Análisis de los experimentos dialélicos a través del procedimiento IML de SAS. *Comunicaciones en Estadística y Cómputo*, Vol. 10 (2):36.
- Ron P J, J L Ramírez D, R Valdivia B, B M Lozano (1999) Comparación de tipos de variedades de maíz desarrolladas por el INIFAP en la región Centro-Occidente de México. *Agrociencia* 33:267-273.
- Rojas B A (1973) *Design and analysis of diallel crosses*. Department of Experimental Statistics, New Mexico State University. 30 p.

APÉNDICE

La estimación de efectos familiares en experimentos de cruza dialélicas sin el ensayo de las cruza recíprocas, se puede lograr a través del algoritmo computacional que sigue, escrito en comandos SAS/IML. Los diseños dos y cuatro de Griffing, ocurren como un caso particular de la clase de experimentos dialélicos que aquí se consideran, pudiendo estimarse efectos familiares también, para ellos.

Archivo de datos

Es indispensable, para el uso del algoritmo computacional, que la información colectada de un experimento dialélico parcial, se organice en un archivo SAS de acuerdo con la estructura y orden siguiente:

```

OPTIONS PS=60 PAGENO=1 NODATE;
DATA MASTACHE;
INPUT CRUZA I J REP Y1 Y2 Y3 Y4; DIALE-
LO=CRUZA;
IF I=1 THEN IG=1;
IF I=2 THEN IG=1;
IF I=3 THEN IG=2;
IF I=4 THEN IG=2;
IF I=5 THEN IG=2;
IF I=6 THEN IG=2;
IF J=1 THEN JG=1;
IF J=2 THEN JG=1;
IF J=3 THEN JG=2;
IF J=4 THEN JG=2;
IF J=5 THEN JG=2;
IF J=6 THEN JG=2;
ANGEL=2;
CARDS;

```

DATOS

Y1, Y2, Y3 y Y4 en el comando INPUT, son las variables por analizar. Obviamente, si sólo se requiere analizar dos variables, Y3 y Y4 no aparecerían en el comando INPUT.

Solamente cabe mencionar que las variables IG y JG son muy importantes, ya que con su empleo se identifica a los miembros de cada familia por medio de un número entero dado por el usuario.

PROGRAMA

TITLE " ANÁLISIS FAMILIAR DE EXPERIMENTOS SIMÉTRICOS BALANCEADOS DE CRUZAS DIALÉLICAS ";

PROC IML;USE MASTACHE; READ ALL INTO MATRIZ;

CRUZA = MATRIZ[,1];
 A = MATRIZ[,2]; B = MATRIZ[,3]; REP = MATRIZ[,5];
 N = NROW(MATRIZ);
 UNO =
 J(N,1,1); CERO = J(N,1,0); MDIS = DESIGN(CRUZA);
 M00 = UNO*INV(UNO`*UNO)*UNO`;
 X = UNO || MDIS; XX = X`*X; XXIG = GINV(XX);
 M = X*XXIG*X`;
 BLOQ = DESIGN(REP);
 W = X || BLOQ; WW = W`*W; WINV = GINV(WW);
 WWW = W*WINV*W`; IDEN = I(N);
 T = NCOL(MDIS); R = MAX(REP); P = MAX(B);
 NC = NCOL(MATRIZ); J = UNO; ANGEL = MATRIZ[,NC];

IF ANY (A=B) THEN Q=1; ELSE Q=0;
 IF ANY (ANGEL=1) THEN PRINT "DISEÑO COMPLETAMENTE AL AZAR"; ELSE
 PRINT "DISEÑO DE BLOQUES AL AZAR";

TITLE " LA MATRIZ DISEÑO Zp ";

A0 = J(N,P,.); B0 = J(N,P,.);
 DO AAA = 1 TO P;
 DO CCC = 1 TO N;
 IF A[CCC,1] = AAA THEN
 A0[CCC,AAA] = 1; ELSE A0[CCC,AAA] = 0;
 IF B[CCC,1] = AAA THEN
 B0[CCC,AAA] = 1; ELSE B0[CCC,AAA] = 0;
 END;
 END;
 AB = A0 + B0; Zp = AB; ZpZp =
 Zp`*Zp; DIAGO = DIAG(ZpZp); Zii = DIAGO*J(P,1,1);
 X0 =
 UNO || AB; X0X0 = X0`*X0; X0IG = GINV(X0X0);
 M0 = X0*X0IG*X0`;

TITLE " MATRICES DISEÑO PARA FAMILIAS (ACG) ";

IG = MATRIZ[,NC-2]; JG = MATRIZ[,NC-1]; FAMILIAS = MAX(JG);

A000 = J(N,FAMILIAS,.); B000 = J(N,FAMILIAS,.);
 DO AAA = 1 TO FAMILIAS;

DO CCC = 1 TO N;
 IF IG[CCC,1] = AAA THEN
 A000[CCC,AAA] = 1; ELSE A000[CCC,AAA] = 0;
 IF JG[CCC,1] = AAA THEN
 B000[CCC,AAA] = 1; ELSE B000[CCC,AAA] = 0;
 END;
 END;
 FAMGDIS = A000 + B000;
 MIG =
 FAMGDIS*GINV(FAMGDIS`*FAMGDIS)*FAMGDIS`;
 COEFIC =
 UNO`*FAMGDIS; COEFIC = DIAG(COEFIC);
 FAMCOEF =
 TRACE(COEFIC); FAMCOEF = COEFIC*P/FAMCOEF;

TITLE " DISEÑOS SIN EVAL. DE C. RECÍPROCAS ";

IF ALL(B >= A) THEN DO; S = MDIS*MDIS`;
 TRA1 = TRACE((M-M0)*S);
 TRA2 = TRACE((M0-M00)*S);
 TRA3 = TRACE((IDEN -M00)*Zp*Zp`);
 QQ = (P-1)*(4*Q+P-2)*R-TRA3;
 IF QQ > 0.1 & ANY(Zii-Zii[1,1]) > 0.1 THEN
 PRINT "DIALÉLICO PARCIAL ASIMÉTRICO SIN C. RECÍPROCAS";
 ELSE IF QQ > 0.1 THEN PRINT "DIALÉLICO PARCIAL SIMÉTRICO-SIN C. RECÍPROCAS";
 ELSE IF Q = 1 THEN PRINT "DISEÑO 2 DE GRIFFING"; ELSE PRINT "DISEÑO 4 DE GRIFFING";
 PRINT N T R P;

TITLE " ANÁLISIS DE VARIANZA ";

GLt = ROUND(TRACE(GINV((M-M00)*(M-M00))));
 IF ANY(ANGEL=1) THEN
 DO; GLr = .; GLe = N-1-GLt; END;
 IF ANY(ANGEL=2) THEN
 DO; GLr = ROUND(TRACE(GINV((WWW-M)*(WWW-M))));
 GLe = N-1-GLr-GLt; END;
 GLacg = ROUND(TRACE(GINV((M0-M00)*(M0-M00)))); GLace = GLt-GLacg;
 FV = J(8,5,.); UN = J(P,1,1);
 PROG = J(P,4,.); PPP = J(P,1,.);
 FAMILIA = J(FAMILIAS,2,.);
 FAMIL = J(FAMILIAS,1,.);
 DO LLL = 1 TO P BY
 1; PPP[LLL,1] = LLL; END;
 DO LLLL = 1 TO FAMILIAS BY 1;
 FAMIL[LLLL,1] = LLLL; END;

DO F = 6 TO (NC-3) BY 1;
 VARIABLE = F-5;
 Y = MATRIZ[,F];


```

FC =
Y`*M00*Y;MEDIA = UNO`*Y/N;SCTOT = Y`*Y-FC;
  IF ANY(ANGEL=1) THEN
DO;SCE=Y`*(IDEN-M)*Y;
CME=SCE/GLe;SCB=.;CMB=.;FBLOQ=.;END;
  IF ANY(ANGEL=2) THEN
DO;SCE=Y`*(IDEN-WWW)*Y;CME=SCE/GLe;
  SCB=Y`*(WWW-M)*Y;
CMB=SCB/GLr;FBLOQ=CMB/CME;END;
  SCCRUZA = (Y`*M*Y) -FC;
CMCRUZA=SCCRUZA/GLt;FCRUZA
=CMCRUZA/CME;
  SCACG = (Y`*M0*Y) -FC;
CMACG=SCACG/GLacg;
CV=(CME**.5)*100/MEDIA;
  SCACE = SCCRUZA-SCACG;
CMACE=SCACE/(GLace);FACE=CMACE/CME;
FACG=CMACG/CMACE;
  GLFAMg =
ROUND(TRACE(GINV((MIG-M00)*(MIG -
M00)))));GLIFAMg=GLacg-GLFAMg;
  SCFAMg = Y`*MIG*Y-FC;CMFAMg
=SCFAMg/GLFAMg; FFAMg =CMFAMg/CMACE;
  SCIFAMg= SCACG-SCFAMg; CMI-
FAMg=SCIFAMg/GLIFAMg; FIFAMg
=CMIFAMg/CMACE;
  PROBFAMg= 1
-PROBF(FFAMg,GLFAMg,GLace);PROBIFMg= 1
-PROBF(FIFAMg,GLIFAMg,GLace);

  PROBB = 1
-PROBF(Fbloq,GLr,GLe);PROBT=1
-PROBF(Fcruza,GLt,GLE);
  PROBACg=1
-PROBF(Facg,GLacg,GLace);PROBACE=1
-PROBF(Face,GLace,GLE);
  FV[1,1] =
GLr;FV[2,1]=GLt;FV[3,1]=GLacg;FV[4,1]=GLFAMg;
FV[5,1]=GLIFAMg;FV[6,1]=GLace;
  FV[7,1]=GLE;FV[8,1]=N-1;
  FV[1,2] =
SCB;FV[2,2]=SCCRUZA;FV[3,2]=SCACG;
FV[4,2]=SCFAMg;FV[5,2]=SCIFAMg;

FV[6,2]=SCace;FV[7,2]=SCe;FV[8,2]=SCTOT;
  FV[1,3] =
CMB;FV[2,3]=CMCRUZA;FV[3,3]=CMACG;FV[4,3]
=CMFAMg;FV[5,3]=CMIFAMg;
  FV[6,3]=CMACE;FV[7,3]=CME;
  FV[1,4] =
FBLOQ;FV[2,4]=FCRUZA;FV[3,4]=FACG;
FV[4,4]=FFAMg;FV[5,4]=FIFAMg;FV[6,4]=FACE;
  FV[1,5] =
PROBB;FV[2,5]=PROBT;FV[3,5]=PROBACg;

```

```

FV[4,5]=PROBFAMg;FV[5,5]=PROBIFMg;
  FV[6,5]=PROBACE;
  CCC = {" GL " "SC" "CM" "F" "Pr >
F"};
  IF ANY(ANGEL=1) THEN DDD={" . "
"CRUZAS" " ACG" " E.FAM." " E.IFAM."
" ACE" "ERROR"
"TOTAL"};
  ELSE DDD={"BLOQUES"
"CRUZAS" " ACG" " E.FAM." " E.IFAM."
" ACE" "ERROR"
"TOTAL"};

  TITLE " ESTIMACIÓN DE LAS COMPONENTES
DE VARIANZA ";
  VARE = CME;VARs=GLace*(CMACE
-CME)/TRA1;
  VARg = ((GLacg)*(CMACG-CME)
-VARs*TRA2)/TRA3;
  IF VARs > 0 THEN VARs=VARs;ELSE
VARs=0;IF VARg>0 THEN VARg=VARg;ELSE
VARg=0;
  RR = (VARs/VARE)*S + I(N);
RR=GINV(RR);

  TITLE " EMC, EMCG Y EL MPLI DE ACG ";

  MU = INV(J`*GRR*J)*J`*GRR*Y;
  EMC = GINV(Zp`*Zp)*Zp`*(Y-MEDIA*J);
MCG=EMC;EMCGMED=EMCG+MU*UN;
  IF VARg > 0 THEN
INVGP=(VARE/VARg)*I(P);ELSE INVGP=0*I(P);
  MPLI =
GINV(Zp`*GRR*Zp+INVGP)*Zp`*GRR*(Y-MU*J);
PLIMED=MPLI+MU*UN;
  PROG[,1]=
EMCG;PROG[,2]=MPLI;PROG[,3]=EMCGMED;
PROG[,4]=MPLIMED;
  EEE = {"EMCG" "MPLI" "EMCG+MU"
"MPLI+MU"};FFF=CHAR(PPP,3,0);

  TITLE " LOS EMCG y LOS MPLI PARA
FAMILIAS ";

  IF VARg > 0 THEN
INVGpp=(VARE/VARg)*FAMCOEF;ELSE
INVGpp=0*FAMCOEF;

PFAMg=GINV(FAMGDIS`*GRR*FAMGDIS
+INVGpp)*FAMGDIS`*GRR*(Y-MU*J);

EFAMg=GINV(FAMGDIS`*GRR*FAMGDIS)
*FAMGDIS`*GRR*(Y-MU*J);

```

```

FAMILIA[,1]=
EFAMg;FAMILIA[,2]=PFAMg;
EEEE={"EMCGg"
"MPLIg"};FFFF=CHAR(FAMIL,3,0);

TITLE " LA MATRIZ DE COEFICIENTES: C DE
ACG ";
CC1=(UNO`*GRR*UNO)| |(UNO`*GRR*Zp);
CC2 =
(UNO`*GRR*Zp)`| |(Zp`*GRR*Zp)+INVGp);
CC3=CC1`| |CC2`;
CCCC =
GINV(CC3);GAMA=2:P+1;CCCC=CCCC[GAMA,
GAMA];

TITLE " IMPRESIÓN DE RESULTADOS ";
PRINT VARIABLE;
PRINT "CUADRO 1. ANÁLISIS DE
VARIANZA.";PRINT FV[ROWNAME=DDD
COLNAME=CCC];
PRINT MEDIA[FORMAT=12.5]
CV[FORMAT=12.5];PRINT ,;
PRINT " ESTIMACIÓN DE LAS COMPO-
NENTES DE VARIANZA";

```

```

PRINT VARE[FORMAT=12.5]
VARs[FORMAT=12.5] VARg[FORMAT=12.5];
PRINT /;
PRINT /;PRINT MU[FORMAT= 12.5];
PRINT ,;
PRINT "CUADRO 2. ESTIMACIÓN Y
PREDICCIÓN DE ACG.";
PRINT PROG[ROWNAME=FFF
COLNAME=EEE FORMAT=12.5];
PRINT ,; PRINT "CUADRO 3. ESTIMACIÓN
Y PREDICCIÓN FAMILIAL.";
PRINT FAMILIA[ROWNAME=FFFF
COLNAME=EEEE FORMAT=12.5];PRINT /;
PRINT "CUADRO 4. LA MATRIZ DE
COEFICIENTES C22 DE ACG. ";
PRINT CCCC[FORMAT=7.3];PRINT /;
END;
END;
QUIT;

```