

Código de SAS para analizar un dialelico completo y heterosis. Un ambiente

Delfina de Jesús Pérez López
Claudia Saavedra Guevara
Martín Rubí Arriaga
J. Ramón Pascual Franco Martínez
Francisco Gutiérrez Rodríguez
Andrés González Huerta[§]

Centro de Investigación y Estudios Avanzados en Fitomeroamiento-Facultad de Ciencias Agrícolas-Universidad Autónoma del Estado de México-Campus Universitario 'El Cerrillo'. El Cerrillo Piedras Blancas, Toluca, Estado de México, México. AP. 435. Tel. 722 2965531, ext. 148. (djperezl@uaemex.mx; csg1003@yahoo.com; m.rubi65@yahoo.com.mx; jrfrancom@uaemex.mx; fgfca@hotmail.com.

[§]Autor para correspondencia: agonzalezh@uaemex.mx.

Resumen

La elaboración de programas para sistema para análisis estadístico (SAS) y su validación con softwares disponibles gratuitamente es indispensable cuando no existen recursos económicos para adquirir la licencia de un paquete estadístico apropiado. En este estudio se presenta un código para SAS y se realiza su validación con el programa propuesto por Zhang y Kang (1997), modificado por Saavedra (2019). El código genera un análisis de varianza con partición de los efectos de tratamientos en progenitores (P), cruza directa (CD), cruza recíproca (CR), P vs cruza, y CD vs CR. Además de generar la comparación de medias de tratamientos con la prueba de Tukey, se estiman los efectos genéticos para progenitores o para sus cruza (Gi, Sij, Rij, Mi); así como, los de heterosis con la media de ambos padres o con el mejor de ellos. Debido a que ambos códigos sólo coinciden en el cálculo de los efectos genéticos previamente indicados, se sugiere su aplicación simultánea para realizar un análisis completo del método 1 de Griffing (1956a, b). El código que ha sido propuesto será de gran utilidad para fitomejoradores y genetistas y especialmente, para estudiantes en ciencias biológicas y agropecuarias de nivel licenciatura y de postgrado con poco entrenamiento en el lenguaje de programación en SAS.

Palabras clave: bloques completos al azar, método 1 de Griffing, modelo 1, prueba de Tukey.

Recibido: febrero de 2020

Aceptado: abril de 2020

Introducción

Las cruzas dialélicas fueron diseñadas antes de los 50's, pero pronto se convirtieron en una poderosa herramienta para mejoradores de plantas y animales, quienes para reconocer el mérito de varios progenitores evaluaron sus progenies a través de los efectos y varianzas de aptitud combinatoria general (ACG) y específica (ACE) (Sprague y Tatum, 1942; Griffing, 1956a, b; González *et al.*, 2007a, b). Con éstas se definen nuevos patrones heteróticos o una población segregante a partir de la cual es posible aislar nuevamente plantas sobresalientes, predecir la respuesta a la selección o el comportamiento de híbridos o sintéticos formados con nuevas líneas (Hallauer y Miranda, 1988; Christie y Shattuck, 1992; González *et al.*, 2007a, b).

El análisis de un experimento de cruzas dialélicas completo sin una computadora personal (PC) es laborioso y para ahorrar tiempo, existen varios paquetes estadísticos como SAS (<https://www.sas.com/store/index.ep>), Excel (Microsoft Office), Indostat (<https://www.indostat.org>), AGD-R (<https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10202>), Agrobase II, generación (<http://www.agronomix.com>), PB Tools (<https://pbtools.software.informer.com/2.0/>), TNAUSTAT (<https://sites.google.com/site/tnaustat>) y GSCA (<https://bioseqdata.com/gsca/gsca.htm>), entre otros; de éstos sólo Agrobase II generación e Indostat deben adquirirse con licencia con un costo superior a \$1 000.0 USD, debido a que se requieren al menos tres módulos para operar adecuadamente ambos softwares. Aunque SAS es el mejor paquete estadístico, es común que los mejoradores y genetistas utilicen varios softwares para analizar los datos de experimentos diseñados en las ciencias agropecuarias y biológicas (Padilla *et al.*, 2019a; Padilla *et al.*, 2019b; Saavedra, 2019).

También, para muchos usuarios es difícil descargar software gratuito debido a que hay problema de incompatibilidad entre éste y sus PC's, existen problemas técnicos durante las descargas, no se obtiene el permiso necesario, los investigadores no responden las solicitudes o el programa no funciona en versiones antiguas o recientes de Windows. En este contexto, sería deseable elaborar y validar algunos códigos para SAS, para versiones 6.01 o superiores (SAS, 1989), que permitan complementar el análisis genético-estadístico para experimentos de cruzas dialélicas completas.

Materiales y métodos

Dialélico completo

En la metodología 1, descrita en Saavedra (2019), el análisis de varianza (Anova) para un solo ambiente contiene repeticiones (R), tratamientos (Trat) y error experimental, su modelo estadístico corresponde a un diseño de bloques completos al azar. En el Anova, los efectos de Trat se dividen en progenitores (P), cruzas directas (CD), cruzas recíprocas (CR), P *vs* cruzas y CD *vs* CR, como lo sugirió González *et al.* (2007b), ambos contrastes estiman heterosis promedio y efectos maternos y no maternos.

El programa calcula las diferencias entre Trat con la prueba de Tukey (SAS, 1989). Este código puede modificarse fácilmente si el usuario requiere de otras pruebas de comparación de medias, o de diversos análisis de regresión y correlación, estos análisis pueden extenderse a series de experimentos en tiempo y espacio (Saavedra, 2019).

En la metodología 2, que corresponde al método 1 de Griffing (1956a, b), el Anova para un solo ensayo tiene repeticiones (R), aptitud combinatoria general (ACG), aptitud combinatoria específica (ACE), efectos maternos (EM) y efectos recíprocos (ER); en las series de experimentos podrían estimarse las interacciones de éstos con sólo dos ambientes.

También, en ambos casos, serían estimables los efectos de g_i para cada progenitor o de s_{ij} para cada cruce, los efectos recíprocos y maternos (Zhang y Kang, 1997). Los componentes de varianza y heredabilidades, y la predicción de híbridos y sintéticos podrían estimarse con otros programas para SAS (Martínez, 1983; González *et al.*, 2007a, b; Montesinos *et al.*, 2007).

Definición de variables en el código

En la base de datos denominada ‘dialelo’ se define female, male, YH, YP, YM, X, Y, A, B, C, D y M, en las variables female y male se indican las combinaciones de cada hembra con cada macho, YH, YP y YM corresponden a las medias de cruce, hembra y macho, respectivamente. En X, Y se capturan los totales para cada pareja de CD y CR. Después de realizado la suma sobre repeticiones en A, B, C, D, cada línea de la CD o de CR aparece dos veces, como hembra y como macho (Y_i. ó Y_i.i.; Y_j. o Y_j.j.)M es la gran media aritmética, GI, SIJ, RIJ y MI son los mismos efectos genéticos que se estiman con las fórmulas propuestas en el método 1 de Griffing (1956a, b).

Valores usados en el código

En este estudio fueron utilizados 96 datos, correspondientes a cuatro progenitores, sus seis cruces directas y sus seis cruces recíprocas, registrados en seis repeticiones (Saavedra, 2019).

Resultados y discusión

Desde su creación en 1972, los programas para SAS para el análisis de experimentos de cruces dialélicas han sido implementados en computadoras personales PC's por varios investigadores. Los grandes logros que se han obtenido para PC's se atribuyen a Schaffer y Usanis (1989); Burow y Coors (1994); Magari y Kang (1994); Zhang y Kang (1997); Martínez (1983, 1991), entre otros. Más recientemente, Mastache y Martínez (1998a, 1998b, 1999a, 1999b), perfeccionaron sus algoritmos para obtener los mejores predictores lineales e insesgados (MPLI) empíricos de los efectos de los progenitores, para ayudar a los usuarios con poco entrenamiento en programación, cuando se utilizan diseños completamente al azar (DCA) y bloques completos al azar (BCA).

También, Mastache y Martínez (2003) obtuvieron un algoritmo integrado para su análisis simultáneo en experimentos balanceados para modelos de efectos fijos o aleatorios. Estos y otros programas podrían emplearse también para validar y para complementar las salidas que se obtuvieron con el código que se propone en el presente estudio (Zhang *et al.*, 2005; Montesinos *et al.*, 2007).

Zhang *et al.* (2005) modificaron los códigos de Zhang y Kang (1997); en Diallel-SAS05, ellos discutieron un programa más eficiente para el análisis genético-estadístico de los cuatro métodos de Griffing (1956a, b), incluyendo los correspondientes a los diseños II y III de Gardner y Eberhart (1966). Este programa es más amigable y es más fácil de modificar que Diallel-SAS, cuando los

progenitores varían de 4 a 12, cuando no hay restricción en el número de ambientes, y cuando se estiman los efectos y varianzas de ACG y ACE para progenitores y cruza, así como de sus interacciones con ambientes. Al igual que con otros paquetes estadísticos, hay problemas para implementarse en computadoras personales con versiones recientes de Windows (Padilla *et al.*, 2019a, b).

Con el programa 1a se calcula el Anova y la comparación de medias (Tukey, $p=0.01$). Como Trat y sus componentes se consideran como efectos fijos, las pruebas de F se prueban con el cuadrado medio del error experimental o residual del modelo. En su código, se utiliza Data, SET, IF-THEN, Anova y GLM para definir subconjuntos de datos. El usuario será cuidadoso en respetar el orden correcto en la base de datos: P, CD y CR, los signos y los coeficientes de los contrastes, al igual que para otros paquetes estadísticos, deben capturarse dentro del programa. Si hay alguna duda para diseñar este tipo de contrastes se sugiere consultar a Padilla *et al.* (2019a).

En la salida del SAS, si $R=6$ y $\text{Trat}=4$, el Anova corresponde a progenitores; su prueba de hipótesis no es correcta, debido a que se construyó como un subconjunto y su cuadrado medio del residual es una fracción de los 96 datos. En este contexto deberá consultarse una tabla de F en esta etapa no hay restricciones con relación al número de variables por analizar. El código puede modificarse para incluir las pruebas de la diferencia mínima significativa (DMS o LSD), Dunnett, o contrastes mutuamente ortogonales, entre otras. Con dos o más variables, es posible modificar el programa para realizar regresión y correlación, estimar estadísticos simples y aplicar metodologías multivariadas, entre otros.

El código correspondiente al programa 1a se presenta a continuación:

```
Data maiz; Input rep trat PVG;Cards; 1 01 758; 1 02 761; 6 15 768; 6 16 758
DATA PADRES;SET MAIZ;IF TRAT>4 THEN DELETE;*sólo progenitores;
DATA CD;SET MAIZ;IF TRAT<5 OR TRAT>10 THEN DELETE;*sólo cruza directas;
DATA CR; SET MAIZ; IF TRAT<11 THEN DELETE; *sólo cruza recíprocas;
PROC ANOVA DATA=MAIZ; CLASS REP TRAT; MODEL PVG=REP TRAT; MEANS
TRAT/TUKEY LINES ALPHA=0.01;*Análisis con los 96 datos;
PROC GLM DATA=MAIZ; CLASS REP TRAT; MODEL PVG=REP TRAT;
CONTRAST "PVSCRUZAS"TRAT 12 12 12 12 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4 -4;
CONTRAST "CD VS CR" TRAT 0 0 0 0 1 1 1 1 1 1 -1 -1 -1 -1 -1;
PROC ANOVA DATA=PADRES; CLASS REP TRAT; MODEL PVG=REP TRAT;*Análisis
de varianza para progenitores;
PROC ANOVA DATA=CD; CLASS REP TRAT; MODEL PVG=REP TRAT;*Análisis de
varianza para cruza directas;
PROC ANOVA DATA=CR; CLASS REP TRAT; MODEL PVG=REP TRAT;*Análisis de
varianza para cruza recíprocas; RUN;
```

Con el programa 1b se obtienen las estimaciones de los efectos genéticos (G_i , S_{ij} , R_{ij} , MI) y la heterosis (%). La definición de variables antes de CARDS debe indicarse correctamente utilizando los datos de los Cuadros 1 y 2 (Saavedra, 2019), pero deben corregirse los valores del denominador de las fórmulas de G_i , S_{ij} , R_{ij} y MI , si cambia R , P o ambos. En esta etapa es fundamental recurrir a los artificios que Martínez (1983) utilizó estableciendo una forma lógica de relacionar las fórmulas de Griffing (1956a, b) con el lenguaje de programación en SAS (SAS Institute, 1989).

Cuadro 1. Peso volumétrico del grano (g L⁻¹) de 16 cruzas formadas con cuatro líneas.

Cruza	R1	R2	R3	R4	R5	R6	Total	Media
1) 1x1	758	734	750	790	758	765	4 555	759.1
2) 2x2	761	762	737	779	763	773	4 575	762.5
3) 3x3	802	812	802	838	793	782	4 829	804.8
4) 4x4	790	768	780	772	783	775	4 668	778
5) 1x2	814	792	770	781	775	755	4 687	781.1
6) 1x3	805	803	806	832	813	824	4 883	813.8
7) 1x4	791	775	777	791	795	780	4 709	784.8
8) 2x3	819	816	793	814	818	786	4 846	807.6
9) 2x4	779	778	758	798	783	755	4 651	775.1
10) 3x4	830	830	850	853	828	806	4 997	832.8
11) 2x1	774	772	786	750	794	769	4 645	774.16
12) 3x1	789	808	816	808	824	806	4 851	808.5
13) 4x1	787	815	815	825	802	796	4 840	806.6
14) 3x2	817	832	808	775	790	797	4 819	803.1
15) 4x2	756	768	756	754	753	768	4 555	759.1
16) 4x3	850	820	840	850	805	758	4 923	820.5
Total	12 722	12 865	12 644	12 810	12 677	12 495	76 033	792.01

Cuadro 2. Valores usados para estimar efectos genéticos y heterosis.

	1	2	3	4	Total
1	4 555	4 687	4 883	4 709	18 834
2	4 645	4 575	4 846	4 651	18 717
3	4 851	4 819	4 829	4 997	19 496
4	4 840	4 555	4 923	4 668	18 986
Total	18 891	18 636	19 481	19 025	76 033

Nota: se sumó sobre repeticiones y los totales de hilera o de columna son la contribución de cada línea hembra o macho, respectivamente.

En algunas columnas, como para la gran media aritmética (M), que es una constante para los 12 apareamientos, hay valores duplicados, pero es fácil establecer a que progenitor o cruza corresponden porque los datos se muestran descendientemente (González *et al.*, 2007a, b; Saavedra, 2019).

El código correspondiente al PROGRAMA 1b se presenta a continuación:

DATA HETERO; INPUT FEMALE MALE YH YP YM X Y A B C D M;

MP= (YP+YM)/2;*para calcular la media de los padres (MP);

BP= MAX (YP, YM);*para elegir al mejor progenitor (BP);

DMP= YH-MP;*para estimar el numerador de la fórmula de heterosis con MP;

HMP= (DMP/MP)*100;*para estimar heterosis con la media de los padres, en %;

DBP= YH-BP;*calcula el numerador de la fórmula de heterosis con BP;

HBP= (DBP/BP)*100;*calcula heterosis con el mejor padre, en %;

GI= (A+B)/48 - M;*estima los efectos de gi;

SIJ= (X+Y)/12 -(A+B+C+D)/48 + M;*calcula los efectos de Sij;

RIJ= (X-Y)/12;*determina los efectos de rij;

MI= (A-B)/48;*calcula los efectos mí;

CARDS;

1 2 781.1 759.1 762.5 4687 4645 18834 18891 18717 18636 792.01

1 3 813.8 759.1 804.8 4883 4851 18834 18891 19496 19481 792.01

1 4 784.8 759.1 778.0 4709 4840 18834 18891 18986 19025 792.01

2 3 807.6 762.5 804.8 4846 4819 18717 18636 19496 19481 792.01

2 4 775.1 762.5 778.0 4651 4555 18717 18636 18986 19025 792.01

3 4 832.8 804.8 778.0 4997 4923 19496 19481 18986 19025 792.01

2 1 774.1 762.5 759.1 4645 4687 18636 18717 18891 18834 792.01

3 1 808.5 804.8 759.1 4851 4883 19481 19496 18891 18834 792.01

4 1 806.6 778.0 759.1 4840 4709 19025 18986 18891 18834 792.01

3 2 803.1 804.8 762.5 4819 4846 19481 19496 18636 18717 792.01

4 2 759.1 778.0 762.5 4555 4651 19025 18986 18636 18717 792.01

4 3 820.5 778.0 804.8 4923 4997 19025 18986 19481 19496 792.01

TITLE 'Efectos de gi, sij,rij, mi y heterosis para el dialélico general';

DATA DOS; SET HETERO; PROC PRINT; RUN;

Procedimiento Anova

Variable dependiente: PVG

Suma de cuadrado de

Fuente DF cuadrados la media F-Valor Pr > F

Modelo 20 51191.16667 2559.55833 9.69 <0.0001

Error 75 19801.82292 264.02431

Total correcto 95 70992.98958

R-cuadrado Coef Var Raiz MSE PVG Media

0.721074 2.051592 16.24882 792.0104

Cuadrado de fuente DF Anova SS la media F-valor Pr > F

rep 5 3365.67708 673.13542 2.55 0.0347

trat 15 47825.48958 3188.36597 12.08 <0.0001

Prueba del rango estudentizado de Tukey (HSD) para PVG

Nota: este test controla el índice de la probabilidad de cometer error experimentwise de tipo I, pero normalmente tiene un índice de error de tipo II más elevado que REGWQ.

Alfa 0.01

Error de grados de libertad 75

Error de cuadrado medio 264.0243

Valor crítico del rango estudentizado 5.76634

Diferencia significativa mínima 38.251

Medias con la misma letra no son significativamente diferentes.

Tukey Agrupamiento Media N trat

A 832.833 6 10

B A 820.500 6 16

B A C 813.833 6 6

B D A C 808.500 6 12

B D A C 807.667 6 8

B D A C 806.667 6 13

B D A C 804.833 6 3

B D A C 803.167 6 14

B D E C 784.833 6 7

D E C 781.167 6 5

D E C 778.000 6 4

D E 775.167 6 9

D E 774.167 6 11

E 762.500 6 2

E 759.167 6 1

E 759.167 6 15

Cuadrado de Contraste DF Contraste SS la media F-Valor Pr > F

P VS CRUZAS 1 8075.086806 8075.086806 30.58 <0.0001

CD VS CR 1 272.222222 272.222222 1.03 0.3132

Variable dependiente: PVG

Suma de Cuadrado de Fuente DF Anova SS la media F-Valor Pr > F

rep 5 1952.875000 390.575000 2.00 0.1376

trat 3 7805.458333 2601.819444 13.30 0.0002

Cuadrado de Fuente DF Anova SS la media F-Valor Pr > F

rep 5 2869.25000 573.85000 3.57 0.0143

trat 5 15157.25000 3031.45000 18.84 <0.0001

Cuadrado de Fuente DF Anova SS la media F-Valor Pr > F

rep 5 1741.80556 348.36111 0.90 0.4948

trat 5 16515.47222 3303.09444 8.56 <0.000

Efectos de GI, SIJ, RIJ, MI y heterosis para la metodología 1

Obs FEMALE MALE YH YP YM X Y A B C D M MP

1 1 2 781.1 759.1 762.5 4687 4645 18834 18891 18717 18636 792.01 760.80

2 1 3 813.8 759.1 804.8 4883 4851 18834 18891 19496 19481 792.01 781.95

3 1 4 784.8 759.1 778.0 4709 4840 18834 18891 18986 19025 792.01 768.55

4 2 3 807.6 762.5 804.8 4846 4819 18717 18636 19496 19481 792.01 783.65

5 2 4 775.1 762.5 778.0 4651 4555 18717 18636 18986 19025 792.01 770.25

6 3 4 832.8 804.8 778.0 4997 4923 19496 19481 18986 19025 792.01 791.40

7 2 1 774.1 762.5 759.1 4645 4687 18636 18717 18891 18834 792.01 760.80

8 3 1 808.5 804.8 759.1 4851 4883 19481 19496 18891 18834 792.01 781.95

9 4 1 806.6 778.0 759.1 4840 4709 19025 18986 18891 18834 792.01 768.55

10 3 2 803.1 804.8 762.5 4819 4846 19481 19496 18636 18717 792.01 783.65

11 4 2 759.1 778.0 762.5 4555 4651 19025 18986 18636 18717 792.01 770.25

12 4 3 820.5 778.0 804.8 4923 4997 19025 18986 19481 19496 792.01 791.40

Obs BP DMP HMP DBP HBP GI SIJ RIJ MI

1 762.5 20.30 2.66824 18.6 2.43934 -6.0725 5.5517 3.5000 -1.1875
 2 804.8 31.85 4.07315 9.0 1.11829 -6.0725 5.2183 2.6667 -1.1875
 3 778.0 16.25 2.11437 6.8 0.87404 -6.0725 9.9267 -10.9167 -1.1875
 4 804.8 23.95 3.05621 2.8 0.34791 -13.8225 7.2183 2.2500 1.6875
 5 778.0 4.85 0.62967 -2.9 -0.37275 -13.8225 -10.9067 8.0000 1.6875
 6 804.8 41.40 5.23124 28.0 3.47913 20.0108 14.7600 6.1667 0.3125
 7 762.5 13.30 1.74816 11.6 1.52131 -13.8225 5.5517 -3.5000 1.6875
 8 804.8 26.55 3.39536 3.7 0.45974 20.0108 5.2183 -2.6667 0.3125
 9 778.0 38.05 4.95088 28.6 3.67609 -0.1142 9.9267 10.9167 -0.8125
 10 804.8 19.45 2.48198 -1.7 -0.21123 20.0108 7.2183 -2.2500 0.3125
 11 778.0 -11.15 -1.44758 -18.9 -2.42931 -0.1142 -10.9067 -8.0000 -0.8125
 12 804.8 29.10 3.67703 15.7 1.95080 -0.1142 14.7600 -6.1667 -0.8125

Los resultados anteriores fueron validados con el programa elaborado por Zhang y Kang (1997). En el Anova el código permite la partición de los efectos de cruza posibles en ACG, ACE, ER y EM, cuando se implementaron SORT, BY, GLM, IF-THEN, DROP, ARRAY, ELSE, GLM, CONTRAST, ESTIMATE y algunas MACROS. En Martínez (1991) se presentan éstas y otras componentes para elaborar el código de referencia.

El programa de Zhang y Kang (1997) aplica a los cuatro métodos de Griffing (1956a, b), para el método 1 se analizan m variables en dos ambientes. En el presente estudio, éste se amoldó a un solo ambiente implementando la restricción IF ENV > 1 THEN DELETE o IF ENV < 2 THEN DELETE, capturado antes de DROP y después del INPUT. Su modificación es más laboriosa para los usuarios con poco entrenamiento en programación y, especialmente cuando el análisis se extiende a series de experimentos (Singh, 1973; Mastache y Martínez, 2003; Zhang *et al.*, 2005).

TNAUSTAT Software, además de calcular los efectos genéticos relacionados con progenitores y con sus cruza en el método 1 de Griffing (1956a, b) también, simultáneamente, permite el cálculo del vigor híbrido con la media de ambos padres, con el mejor de ellos y adicionalmente, con base a heterosis comercial. Éste tiene la ventaja adicional de estimar los parámetros genéticos correspondientes al diseño de apareamiento I, propuesto por Hayman (1954). Sin embargo, este software fue diseñado para funcionar adecuadamente en una plataforma con MS Dos, por lo que anticipadamente debe descargarse del internet DOSBox software.

El Programa de Zhang y Kang (1997), Modificado por Saavedra (2019), se presenta a continuación:

```

OPTIONS PS=56 LS=78; TITLE 'METHOD 1'; DATA METHOD1;
INPUT I J REP HYBRID YIELD ENV; IF ENV>1 THEN DELETE;DROP N NI NJ P;
P=4;*NUMBER OF PARENTAL LINES? ; ARRAY GCA (N) G1 G2 G3;DO N=1 TO (P-1);
GCA= ((I=N)-(I=P)) + ((J=N)-(J=P)); END;ARRAY SCA(N) S11 S12 S13 S22 S23 S33;
N=0; DO NI=1 TO (P-1); DO NJ=NI TO (P-1); N+1; IF NI=NJ THEN DO;
SCA=(I=NI)*((J=NJ)-(J=P))+I=P*((J=P)-(J=NI));END;ELSE DO;
SCA=(I=NI)*(J=NJ)-(J=P)*((I=NI)+(I=NJ)-(I=P)*2)+(I=NJ)*(J=NI)
-(I=P)*((J=NI)+(J=NJ));END;END;END;
ARRAY REC (N) R12 R13 R14 R23 R24 R34; N=0; DO NI=1 TO (P-1);
DO NJ= (NI+1) TO P; N+1; REC= (I=NI)*(J=NJ)-(j=NI)*(I=NJ); END;END;
ARRAY MAT (N) M1 M2 M3; DO N=1 TO (P-1); MAT= (I=N) + (J=P)-(J=N)-(I=P);

```



```

END;ARRAY NONM (N) N12 N13 N23;N=0;DO NI=1 TO (P-2);DO NJ=(NI+1) TO (P-
1);N+1;NONM=((I=NI)*(J=NJ))-(I=NJ)*(J=NI)-((I=NI)*(J=P))+(I=NJ)*(J=P)
+ ((I=P)*((J=NI)-(J=NJ))); END;END;CARDS;
1 1 1 01 758 1
1 2 1 02 814 1
1 3 1 03 805 1
4 3 6 15 758 1
4 4 6 16 775 1
PROC SORT; BY REP ENV I J; PROC GLM; CLASS REP ENV HYBRID; MODEL
YIELD=ENV REP(ENV) HYBRID HYBRID*ENV; TEST H=HYBRID
E=HYBRID*ENV;LSMEANS HYBRID;
RUN; TITLE 'DIALLEL-SAS 1'; PROC GLM; CLASS REP ENV HYBRID;
MODEL YIELD= ENV REP (ENV) G1 G2 G3 S11 S12 S13 S22 S23 S33 R12 R13 R14 R23
R24 R34 G1*ENV G2*ENV G3*ENV S11*ENV S12*ENV S13*ENV S22*ENV S23*ENV
S33*ENV R12*ENV R13*ENV R14*ENV R23*ENV R24*ENV R34*ENV;
%MACRO GCASCA; CONTRAST 'GCA' G1 1, G2 1, G3 1;
CONTRAST 'SCA' S11 1, S12 1, S13 1, S22 1, S23 1, S33 1;
ESTIMATE 'G1' G1 1; ESTIMATE 'G2' G2 1; ESTIMATE 'G3' G3 1;
Estimate 'G4' G1 -1 G2 -1 G3 -1;
ESTIMATE 'S11' S11 1; ESTIMATE 'S12' S12 1; ESTIMATE 'S13' S13 1;
ESTIMATE 'S22' S22 1; ESTIMATE 'S23' S23 1; ESTIMATE 'S33' S33 1;
Estimate 'S14' S11 -1 S12 -1 S13 -1;
Estimate 'S24' S12 -1 S22 -1 S23 -1;
Estimate 'S34' S13 -1 S23 -1 S33 -1;
Estimate 'S44' S11 1 S12 2 S13 2 S22 1 S23 2 S33 1;
%MEND GCASCA; %GCASCA %MACRO INTERACT;
CONTRAST 'GCA*ENV' G1*ENV 1 -1, G2*ENV 1 -1, G3*ENV 1 -1;
CONTRAST 'SCA*ENV' S11*ENV 1 -1, S12*ENV 1 -1, S13*ENV 1 -1, S22*ENV 1 -1,
S23*ENV 1 -1, S33*ENV 1 -1; %MEND INTERACT; %INTERACT
CONTRAST 'REC' R12 1, R13 1, R14 1, R23 1, R24 1, R34 1;
ESTIMATE 'R12' R12 1; ESTIMATE 'R13' R13 1; ESTIMATE 'R14' R14 1;
ESTIMATE 'R23' R23 1; Estimate 'R24' R24 1; ESTIMATE 'R34' R34 1;
CONTRAST 'REC*ENV' R12*ENV 1 -1,R13*ENV 1 -1,R14*ENV 1 -1,R23*ENV 1 -
1,R24*ENV 1 -1,R34*ENV 1 -1;
CONTRAST 'MAT SS' R12 1 R13 1 R14 1, R12 -1 R23 1 R24 1, R13 -1 R23 -1 R34 1, R14 -1
R24 -1 R34 -1; ESTIMATE 'MAT1' R12 1 R13 1 R14 1/DIVISOR=3;
ESTIMATE 'MAT2' R12 -1 R23 1 R24 1/DIVISOR=3;
ESTIMATE 'MAT3' R13 -1 R23 -1 R34 1/DIVISOR=3;
ESTIMATE 'MAT4' R14 -1 R24 -1 R34 -1/DIVISOR=3; RUN;
TITLE 'DIALLEL-SAS 2'; PROC GLM; CLASS REP ENV HYBRID;
MODEL YIELD= ENV REP (ENV) G1 G2 G3 S11 S12 S13 S22 S23 S33
M1 M2 M3 N12 N13 N23 G1*ENV G2*ENV G3*ENV
S11*ENV S12*ENV S13*ENV S22*ENV S23*ENV S33*ENV
M1*ENV M2*ENV M3*ENV N12*ENV N13*ENV N23*ENV;
%GCASCA %INTERACT
CONTRAST 'MAT SS' M1 1, M2 1, M3 1;
CONTRAST 'NONM SS' N12 1, N13 1, N23 1;

```

```

CONTRAST 'MAT*ENV' M1*ENV 1 -1, M2*ENV 1 -1, M3*ENV 1 -1;
CONTRAST 'NONM*ENV' N12*ENV 1 -1, N13*ENV 1 -1, N23*ENV 1 -1;
ESTIMATE 'M1' M1 1; ESTIMATE 'M2' M2 1; ESTIMATE 'M3' M3 1;
Estimate 'M4' M1 -1 M2 -1 M3 -1;
ESTIMATE 'N12' N12 1; ESTIMATE 'N13' N13 1; ESTIMATE 'N23' N23 1;
Estimate 'N14' N12 -1 N13 -1;
Estimate 'N24' N12 1 N23 -1;
Estimate 'N34' N13 1 N23 1; RUN;

```

Para validar el código que se presenta en el programa 1a, se muestran algunos resultados que genera el programa de Zhang y Kang (1997).

Procedimiento GLM

Variable dependiente: YIELD

Suma de Cuadrado de

Fuente DF cuadrados la media F-Valor Pr > F

Modelo 20 51191.16667 2559.55833 9.69 <.0001

Error 75 19801.82292 264.02431

Total correcto 95 70992.98958

R-cuadrado Coef Var Raiz MSE YIELD Media

0.721074 2.051592 16.24882 792.0104

Cuadrado de Fuente DF Tipo I SS la media F-Valor Pr > F

ENV 0 0.00000

REP (ENV) 5 3365.67708 673.13542 2.55 0.0347

HYBRID 15 47825.48958 3188.36597 12.08 <.0001

ENV*HYBRID 0 0.00000.

Cuadrado de Contraste DF Contraste SS la media F-Valor Pr > F

GCA 3 30162.39583 10054.13194 38.08 <.0001

SCA 6 14715.59375 2452.59896 9.29 <.0001

REC 6 2947.50000 491.25000 1.86 0.0988

(MAT SS) (3) (240.75000) 80.25000 0.30 0.8224

(NONM SS) (3) (2706.75000) 902.25000 3.42 0.0216

Error Parámetro Estimación estándar Valor t Pr > |t|

G1 -6.0729167 2.03110309 -2.99 0.0038

G2 -13.8229167 2.03110309 -6.81 <.0001

G3 20.0104167 2.03110309 9.85 <.0001

G4 -0.1145833 2.03110309 -0.06 0.9552

S12 5.5520833 3.70826994 1.50 0.1385

S13 5.2187500 3.70826994 1.41 0.1635

S23 7.2187500 3.70826994 1.95 0.0553

S14 9.9270833 3.70826994 2.68 0.0091

S24 -10.9062500 3.70826994 -2.94 0.0043

S34 14.7604167 3.70826994 3.98 0.0002

M1 -1.1875000 2.03110309 -0.58 0.5605

M2 1.6875000 2.03110309 0.83 0.4087

M3 0.3125000 2.03110309 0.15 0.8781

M4 -0.8125000 2.03110309 -0.40 0.6903

R12 3.5000000 4.69063167 0.75 0.4579

R13 2.6666667 4.69063167 0.57 0.5714
R14 -10.9166667 4.69063167 -2.33 0.0226
R23 2.2500000 4.69063167 0.48 0.6329
R24 8.0000000 4.69063167 1.71 0.0922
R34 6.1666667 4.69063167 1.31 0.1926

Conclusiones

Los programas '1a' y '1b' son fáciles de utilizar y modificar para realizar un análisis de varianza en un solo ambiente, con la subdivisión de los efectos de los tratamientos en progenitores (P), cruza directa (CD), cruza recíproca (CR), P *versus* cruza y CD *versus* CR. También es útil para realizar la comparación de medias de tratamientos (Tukey, $p=0.01$) y para estimar heterosis con la media de los padres y con el mejor de éstos cuando se analiza una variable.

El programa de Zhang y Kang (1997) fue diseñado para analizar 'm' variables, pero es más difícil de manipular cuando progenitores y ambientes es diferente de 5 y 2, respectivamente. Por esta restricción, fue necesario modificar el código con $P=4$. El código propuesto por Zhang y Kang estima aptitud combinatoria general y específica, efectos recíprocos y maternos, pero no incluye la prueba de Tukey ni la estimación de heterosis.

Los tres códigos para SAS corren en versiones de más de 10 años de su liberación comercial y en las académicas de prueba más recientes. El programa de Zhang y Kang (1997) permitió la validación confiable del código propuesto en el presente estudio cuando se estimaron los efectos genéticos, pero los tres códigos deben utilizarse para realizar un análisis dialélico más completo.

Literatura citada

- Burow, M. D. and Coors, J. G. 1994. DIALLEL: a microcomputer program for the simulation and analysis of diallel crosses. *Agron. J.* 86(1):154-158.
- Christie, B. R. and Shattuck, V. I. 1992. The diallel cross: design, analysis and use for plant breeders. *Plant Breeding Reviews.* 9(1):9-36.
- Gardner, C. O. and Eberhart, S. A. 1966. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics.* 22(3): 439-452.
- González, H. A.; Sahagún, C. J. y Pérez, L. D. J. 2007a. Estudio de ocho líneas de maíz en un experimento dialélico incompleto. *Ciencias Agrícolas Informa.* 16(1):3-9.
- González, H. A.; Pérez, L. D.; Sahagún, C. J.; Norman, M. T. H.; Balbuena, M. A. and Gutiérrez, R. F. 2007b. Análisis de una crusa dialélica completa de líneas endogámicas de maíz. *Ciencias Agrícolas Informa.* 16(1):10-17.
- Griffing, B. 1956a. A generalized treatment of the use of diallel crosses in quantitative inheritance. *Heredity.* 10(1):31-50.
- Griffing, B. 1956b. Concept of general and specific combining ability in relation to diallel crossing systems. *Austr. J. Biol. Sci.* 9(4):463-491.
- Hallauer, A. R.; Miranda, F. O. J. B. 1988. *Quantitative Genetics in Maize Breeding.* Iowa State University Press, Ames. Second Edition. USA. 468 p.
- Hayman, B. I. 1954. The theory and analysis of the diallel crosses. *Genetics.* 39(1):798-809.

- Magari, R. and Kang, M. S. 1994. Interactive BASIC program for Griffing's diallel analysis. *Journal of Heredity*. 85(4):336.
- Martínez, G. A. 1983. Diseño y análisis de los experimentos de cruza dialélicas. Centro de Estadística y Cálculo. Colegio de Postgraduados. Chapingo, Estado de México. 252 p.
- Martínez, G. A. 1991. Análisis de los experimentos dialélicos a través del procedimiento IML del SAS. *Comunicaciones en Estadística y Cómputo*. 10(2):1-36.
- Mastache, L. A. A.; Martínez, G. A.; Castillo, M. A. y González, C. F. V. 1998a. Los mejores predictores lineales e insesgados (MPLI) en experimentos dialélicos parciales sin efectos maternos. *Revista Fitotecnia Mexicana*. 21(1):49-60.
- Mastache, L. A. A.; Martínez, G. A. y Castillo, M. A. 1998b. Los mejores predictores lineales e insesgados (MPLI) en experimentos dialélicos parciales con efectos maternos. *Revista Fitotecnia Mexicana*. 21(2):171-184.
- Mastache, L. A. A.; Martínez, G. A. y Castillo, M. A. 1999a. Los mejores predictores lineales e insesgados (MPLI) en los diseños dos y cuatro de Griffing. *Agrociencia*. 33(1):81-90.
- Mastache, L. A. A.; Martínez, G. A. y Castillo, M. A. 1999b. Los mejores predictores lineales e insesgados (MPLI) en los diseños uno y tres de Griffing. *Agrociencia*. 33(3):349-359.
- Mastache, L. A. A. y Martínez, G. A. 2003. Un algoritmo para el análisis, estimación y predicción en experimentos dialélicos balanceados. *Revista Fitotecnia Mexicana*. 26(3):191-200.
- Montesinos, L. O. A.; Mastache, L. A. A.; Luna, E. I. y Hidalgo, C. J. V. 2007. Mejor predictor lineal e insesgado combinado para aptitud combinatoria general y análisis combinado de los diseños uno y tres de Griffing. *Técnica Pecuaria en México*. 45(2):131-146.
- Padilla, L. A.; González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R. F. y Franco, M. J. R. P. 2019a. InfoStat, InfoGen y SAS para contrastes mutuamente ortogonales en experimentos en bloques completos al azar en parcelas subdivididas. *Rev. Mex. Cienc. Agríc.* 10(6):1417-1431.
- Padilla, L. A.; González, H. A.; Pérez, L. D. J.; Rubí, A. M.; Gutiérrez, R. F.; Ramírez, D. J. F.; Franco, M. J. R. P. y Serrato, C. R. 2019b. Programas para SAS e InfoStat para analizar una serie de experimentos en parcelas subdivididas. En: temas selectos en la innovación de las ciencias agropecuarias. Alfaomega Grupo Editor SA. de CV. Primera edición (Salgado y otros, eds.). México, DF. 724 p. ISBN: 9786075384115.
- SAS Institute, Inc. 1989. SAS/IML software: Usage and reference. Version 6. First edition. Cary, N. C.
- Saavedra, G. C. 2019. Estimación de parámetros genéticos en maíz con dos metodologías usando datos de una cruce dialélica completa. I. Un ambiente. Tesis de Maestro en Fitomejoramiento. Facultad de Ciencias Agrícolas, Universidad Autónoma del Estado de México. Toluca, Estado de México. 96 p.
- Singh, D. 1973. Diallel analysis for combining ability over several environments-II. *Indian Journal of Genetics and Plant Breeding*. 33(1):469-481.
- Schaffer, H. E. y Usanis, R. A. 1989. General least squares analysis of diallel experiments: A computer program. *Genetics Dep. Res. Rep.* 1. North Carolina State University, Raleigh. 61 p.
- Sprague, G. F. and Tatum, L. A. 1942. General vs specific combining ability in single crosses of corn. *J. Amer. Soc. Agron.* 34(1):923-932.
- Zhang, Y. and Kang, M. S. 1997. DIALLEL-SAS: A SAS program for Griffing's Diallel Analyses. *Agronomy Journal*. 89(2):176-182.
- Zhang, Y.; Kang, M. S. and Lamkey, K. R. 2005. DIALLEL-SAS05: A comprehensive program for Griffing's and Gardner-Eberhart analysis. *Agron. J.* 97(4):1097-1106.