DISEÑO Y ANALISIS DE CRUZAS DIALELICAS

BASILIO A. ROJAS

Centro de Estadística y Cálculo, Colegio de Postgraduados, Chapingo, Méx.

Sinopsis

En el presente estudio se describe el método para formular un diseño dialético, completo o parcial, como para su correspondiente análisis.

Un diseño dialélico está caracterizado por el número de líneas n y por el número b. b= 2, 3,..... hasta (n-1)/2. El número de cruzas que incluye el diseño es bn y cada línea se cruza con 2b líneas diferentes. Habrá tantos dialélicos parciales como números enteros hay entre 1 y (n-1)/2, sin incluir a los dos límites 1 y (n-1)/2. El dialélico parcial mínimo ocurre cuando b=2.

Summary

The method to formulate complete or partial diallel models and the methods to analyze them are described in this study.

A diallel model is characterized by the number of lines n an by the number b. b=2,3,... until (n-1)/2. The number of crosses in the model is bn. Each line is crossed with 2b different lines. There will be as many partial diallels as the integral numbers between 1 and (n-1)/2, excluding the two limits 1 and (n-1)/2. The minimum partial diallel happen when b=2.

Introducción

En investigaciones genéticas, especialmente en las concernientes al maíz híbrido, se han utilizado con amplitud las cruzas dialélicas. Estas consisten en probar todas las cruzas posibles entre n líneas con exclusión de las autofecundaciones. El número de cruzas posibles es n(n-1)/2, el cual crece rápidamente con n. Por ejemplo, si n=20 líneas, el número total de cruzas es 10 x 19 = 190. Por esta razón es natural preguntarse si puede elaborarse un diseño tal que pueda proporcionar suficiente información con menor número de cruzas.

El autor elaboró diversos diseños dialélicos parciales en 1962 para satisfacer la inquietud de diversos genetistas mexicanos. Al tratar este problema en ese entonces y encontrar una solución relativamente fácil y de interesantes implicaciones para los genetistas, el autor se preguntó si la resolución estaba ya consignada en la literatura científica. Para investigar ésto se solicitó la colaboración del Dr. Mario Gutiérrez G., quien posteriormente indicó la referencia (1) Kemp-

(1) El autor agradece profundamente al Dr. Eduardo Casas el estímulo moral recibido y a ello se debe principalmente haber escrito este artículo.

tohrne and Curnow (1961). Sin embargo, en las investigaciones agronómicas que se han llevado a cabo en México, estos diseños reducidos no se han empleado en la medida deseable. En una conferencia que el autor dio a la Sociedad Mexicana de Genetistas, en febrero de 1968, se refirió nuevamente a los dialélicos pero vistos desde un aspecto más amplio. Se indicó la gran diversidad de diseños útiles en investigaciones científicas que pueden elaborarse a partir de relaciones de simetría gráfica, siendo los diseños dialélicos parte de ese amplio conjunto.

Familia de Diseños Dialélicos

Llamemos (i, j) a la cruza de la línea i con la j, i j. Es obvio que (i, j) = (j, i), y por lo tanto, para no duplicar la inclusión de una misma cruza, sólo nos referimos a:

$$\begin{aligned}
&\text{cruza} = (i, j);\ i = 1, 2 \ldots \\
&i = 1, 2, 3, \ldots\ldots\ldots\ n; \\
&j = i+1,\ i+2,\ \ldots\ i+b\ \text{(módulo } n); \\
&2 \leq b \leq (n-1)/2;\ \text{(b, número entero)}, \qquad (1)
\end{aligned}$$

El diseño dialélico queda determinado por las condiciones (1). Tenemos n líneas y el diseño en particular está caracterizado por el valor de b que es elegido por el investigador. b puede ser 2, 3, ... hasta el número entero próximo menor a (n-1)/2. Cuando b es igual a (n-1)/2, único posible valor fraccionario permitido, tendremos el dialélico completo.

La expresión $j = i+1,\ i+2, \ldots\ i+b$ (módulo n), indica que si el valor $i+r$, (para $r = 1, 2 \ldots b$) es menor o igual a n, j será igual a $i+r$; pero que si $i+r$ es mayor que n, daremos a j el valor de la diferencia $i+r-n$.

El dialélico con características n y b es tal que cualquier línea se cruza con 2b líneas diferentes y tendrá un total de b x n cruzas diferentes. Habrá tantos diseños dialélicos parciales como números enteros entre uno y (n-1)/2, sin incluir a estos dos límites. El dialélico parcial mínimo ocurre cuando b=2. Existen dialélico parciales cuando $n \geq 7$.

Ejemplo: Supongamos n=10. Por lo mismo (n-1)/2 = 9/2=4.5 b, por lo tanto puede tener los valores 2, 3 y 4 y para cada uno de éstos corresponde un diferente dialélico parcial. Si b=2, el número de cruzas diferentes que incluirá el diseño serán 2 x 10 = 20; si b=3, el número de cruzas será 3 x 10 = 30; si b=4, el número a que nos referimos será 10. El dialélico completo incluirá 10 x 9/2 = 45 cruzas.

Para la obtención de las cruzas que deba incluir un diseño utilizamos las condiciones (1). Primeramente damos a (i, j) el valor 1 a i y obtenemos los j correspondientes según el valor de b. Enseguida pasamos a i=2 y obtenemos sus j, y así sucesivamente. Para ilustrar ésto, consideramos n=10 y b=3.

Para i=1	:	(i,j):	(1,1+1),	(1,1+2),	(1,1+3)	equivalentes a
			(1,2),	(1,3),	(1,4)	
Para i=2	:	(i,j):	(2,2+1),	(2,2+2),	(2,2+3)	equivalentes a
			(2,3),	(2,4),	(2,5)	

Para $i=9$: (i,j): $(9,9+1)$, $(9,9+2)$, $(9,9+3)$ equivalentes a $(9,10)$, $(9,1)$, $(9,2)$

Para $i=10$: (i,j): $(10,10+1)$, $(10,10+2)$, $(10,10+3)$ equivalentes a: $(10,1)$, $(10,2)$, $(10,3)$

Las 30 cruzas que requiere el dialélico parcial para $n=10$ y $b=3$ serán entonces: (1,2), (1,3), (1,4), (2,3), (2,4), (2,5), (3,4), (3,5), (3,6), (4,5), (4,6), (4,7), (5,6), (5,7), (5,8), (6,7), (6,8), (6,9), (7,8), (7,9), (7,10), (8,9), (8,10), (8,1), (9,10), (9,1), (9,2), (10,1), (10,2), y (10,3).

Análisis Estadístico

Supongamos un dialélico con características n y b, que incluye nb cruzas distintas, y que éstas, para simplificación de nuestra exposición, se han probado en un experimento de bloques al azar con r repeticiones. El análisis de la varianza correspondiente se muestran en el Cuadro 1.

Cuadro 1. Análisis de la Varianza. Bloques al Azar.

Fuente de Variación	*G.L.*	*S.C.*	*C.M*
Repetición	r-1	R	
Cruzas	nb-1	C	C'
Error Experimental	(nb-1) (r-1)	E	E'
T o t a l	nbr-1	T	

Nuestro propósito es descomponer los nb-1 grados de libertad correspondientes a cruzas en aptitudes combinatorias general y específica, así como sus respectivas sumas de cuadrados, cuadrados medios y las esperanzas matemáticas de éstos. El análisis de la varianza con esta descomposición será el que se indica en el Cuadro 2.

Cuadro 2. Análisis de la Varianza. Diseño Dialélico.

Fuente de Variación	*G.L.*	*S.C.*	*C.M.*	*E.C.M.*
Aptitud Combinatoria General	n-1	L	L'	L"
Aptitud Combinatoria Específica	n(b-1)	S	S'	S"
Error Experimental	(nb-1) (r-1)	E	E'	E"

En el Cuadro 1 se calculan las sumas de cuadrados indicadas sin que ellas revistan ningún problema. Para construir el Cuadro 2 sólo es necesario calcular L, suma de cuadrados para aptitud combinatoria general, ya que la correspondiente a la específica S, se obtiene por diferencia. es decir

$$S = C - L \qquad (2)$$

Suma de Cuadrados L. El análisis de la varianza es básicamente un análisis de regresión sobre el modelo matemático que se supone explica el mecanismo de acción de efectos. Nuestro modelo es:

$$y_{ijk} = \mu + g_i + g_j + s_{ij} + r_k + e_{ij} \ldots$$

en el que y_{ij} es el rendimiento de la cruza (i, j) en la repetición k. g_i y g_j son los efectos aditivos de las líneas i y j respectivamente. s_{ij} es el efecto específico en el cruzamiento de las mismas líneas. r_k es el efecto aditivo de la repetición k. e_{ijk} es el error experimental de la observación (ijk).

El promedio de la cruza (i, j) en las k repeticiones lo llamaremos y_{ij} y será igual a

$$y_{ij} = \mu + g_i + g_j + s_{ij} + e_{ij} \qquad (4)$$

e_{ij} es el error experimental promedio para la cruza (i, j) en las k repeticiones. El modelo (4) se puede escribir:

$$y_{ij} = \mu + g_i + g_j + (s_{ij} + e_{ij}) = \mu + g_i + g_j + \varepsilon_{ij} \qquad (5)$$

o sea que $\varepsilon_{ij} = s_{ij} + e_{ij}$ es la discrepancia de la parte aditiva $\mu + g_i + g_j$. El modelo (5) no es de rango completo para las estimaciones de n + 1 parámetros μ y g_i, i = 1, 2, ... n. Sin embargo mediante la simple reparametrización,

$$G_i = \frac{\mu}{2} + g_i,$$

el modelo (5) queda: (6)

$$y_{ij} = G_i + G_j + \varepsilon_{ij},$$

el cual es de rango completo y por lo mismo la estimación de las G_i se realiza utilizando los conocimientos clásicos de la regresión, que esencialmente consisten en escribir las ecuaciones normales y resolver el sistema de ecuaciones que será en nuestro caso no-singular. Las ecuaciones normales en notación matricial son:

$$Y = AG \qquad (7)$$

en la que Y es un vector columna n x 1, A es una matriz no singular de dimensiones n x n y G es vector columna n x 1.

Tenemos:

$$Y' = \left(\frac{Y_{1..}}{r}, \frac{Y_{2..}}{r}, \;..\; \frac{Y_{i..}}{r}, \;..\; \frac{Y_{n..}}{r} \right) \qquad (8)$$

$$G' = (G_1, G_2, \ldots, G_i, \ldots G_n). \qquad (9)$$

Los elementos de la matriz A son como sigue:

$a_{ii} = 2b$

$a_{ij} = 1$ si la cruza ij sí está incluida en la prueba (10)
$a_{ij} = 0$ si la cruza ij no está incluida en la prueba

La resolución del sistema (6) nos proporciona las estimaciones de G_i:

$$\hat{G} = A^{-1} Y \tag{11}$$

en la que $\hat{G}' = (\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_i, \ldots, \hat{G}_n)$. La suma de cuadrados debida a las G_i es $\sum_{i=1}^{n} \hat{G}_i Y_i$ y la suma de cuadros L es la anterior corregida por el factor de corrección equivalente a $\frac{Y^2\ldots}{4nbr}$. Debe tenerse presente que $Y\ldots$ es la suma de las $Y_i\ldots$ y por lo tanto $Y\ldots$ representa el doble de la suma de las cruzas en sus r repeticiones. Así pues, la suma de cuadrados para aptitud combinatoria general es:

$$L = \sum_{i=1}^{n} \hat{G}_i Y_{i..} - \frac{Y^2\ldots}{4nbr} \tag{12}$$

Calculado el valor de L con la fórmula (12), la suma de cuadrados para aptitud combinatoria específica S, se obtiene por medio de (2). Para completar el cuadro del análisis de la varianza slo resta obtener las esperanzas matemáticas de los cuadros medios.

Esperanza Matemática de los Cuadrados Medios. Como es costumbre, en el modelo (3) se asume que las g_i, s_{ij} y e_{ijk} son variables aleatorias, con valores medios igual a cero, con covarianzas también cero y con varianzas:

$$V(g_i) = \sigma_g^2,\ V(s_{ij}) = \sigma_s^2,\ V(e_{ijk}) = \sigma^2.$$

Para cualquier diseño dialélico de características n y b, las esperanzas de los cuadrados medios del error, σ^2, y de la aptitud combinatoria específica, $\sigma^2 + r\sigma_s^2$, son invariables, como es fácil demostrar. El cuadrado medio de la aptitud combinatoria general es $\sigma^2 + r\sigma_s^2 + t\sigma_g^2$, en la que el coeficiente t depende de b. Se puede demostrar que:

$$(n-1)\ t = 2bn - \frac{\sum_{j=1}^{n} a_{ij} + (n-1) \sum_{j=1}^{n} a_{1j}\, a_{2j}}{4b} \tag{13}$$

La fórmula anterior, aunque de expresión aparentemente complicada, es de cálculo sencillo. $\Sigma a_{1j}\, a_{2j}$ es la suma de los cuadrados de los términos de la primera hilera (podría ser cualquier hilera, pues todas tienen los mismos términos) de la matriz A, definida en (10). Asimismo $\Sigma a_{1j}\, a_{2j}$ es la suma de los productos de los términos de la primera hilera por los respectivos de la segunda hilera (podría ser cualquier par de hileras).

Con lo anterior hemos cubierto los aspectos esenciales que envuelve el diseño y análisis de un diseño dialélico. Enseguida mostraremos el método de simetría gráfica que objetivamente produce el diseño deseado y su matriz A para la elaboración del análisis.

Método de Simetría Gráfica. En realidad, el procedimiento analítico señalado arriba fue el resultado de haber observado que los diseños dialélicos se producen al construir una tabla de n hileras y n columnas e insertándole las cruzas por realizar, pero cumpliéndose con un arreglo simétrico, como se verá enseguida. La ventaja de este procedimiento es que se reproduce directamente la matriz A y con ella la resolución del sistema de ecuaciones para la obtención de las $\hat{G}i$, para lo cual puede seguirse uno de los procedimientos conocidos: o bien se obtiene A^{-1} y se utiliza la ecuación (11). Enseguida se aplica (12) para obtener la suma de cuadrados de aptitud combinatoria general y la (13) para obtener su esperanza matemática.

En la tabla de n hileras y n columnas se numeran aquellas y éstas sucesivamente de 1, 2 hasta n. Se aumenta b columnas mas enseguida de la n y éstas se vuelven a numerar sucesivamente 1, 2...b. En la primera hilera se pone una marca (x) en las columnas 2, 3, etc., tantas columnas sucesivas como b. Enseguida se marcan las celdas de la hilera 2, empezando con la columna 3 y hasta b columnas. Así seguidamente hasta la hilera n. Se observará una hilera en que las marcas rebasan la n-ésima columna y que el conjunto tiene forma de escalera. Las marcas que quedan a la derecha de la n-ésima columna se han indicado con otro símbolo (Φ), para expresar que son "imágenes" de las cruzas. Estas imágenes se marcan con (x) dentro del cuerpo de la tabla n x n, pero invirtiendo la hilera por columna y columna por hilera. Quedarán así dentro de la tabla n x n y arriba de la diagonal las cruzas que requerirá el diseño.

Diagrama 1. Formación de la escalera

		1	2	3	4	5	6	7	8	9	10	1	2	3	
	1		x	x	x										
	2			x	x	x									
	3				x	x	x								
	4					x	x	x							
	5						x	x	x						
	6							x	x	x					
	7								x	x	x				
	8									x	x	Φ			
	9										x	Φ	Φ		
	10											Φ	Φ	Φ	

Diagrama 2. Inserción de las imágenes (Φ) en la tabla de cruzas

	1	2	3	4	5	6	7	8	9	10
1		x	x	x				x	x	x
2			x	x	x				x	x
3				x	x	x				x
4					x	x	x			
5						x	x	x		
6							x	x	x	
7								x	x	x
8									x	x
9										x
10										

Diagrama 3. Inserción de imágenes dentro de la tabla de cruzas

	1	2	3	4	5	6	7	8	9	10
1		x	x	x				x	x	x
2	Φ		x	x	x				x	x
3	Φ	Φ		x	x	x				x
4	Φ	Φ	Φ		x	x	x			
5		Φ	Φ	Φ		x	x	x		
6			Φ	Φ	Φ		x	x	x	
7				Φ	Φ	Φ		x	x	x
8	Φ				Φ	Φ	Φ		x	x
9	Φ	Φ				Φ	Φ	Φ		x
10	Φ	Φ	Φ				Φ	Φ	Φ	

Para la construcción de la matriz A correspondiente al dialélico n y b, únicamente se marcan en la tabla n x n. las imágenes de las cruzas, por ejemplo. cruza (i, j) = cruza (j, i). La matriz A se forma poniendo 1 en la celda si ella está marcada y 0 si no está marcada. Además en todas las celdas de la diagonal (i, i) se pone 2b.

Las explicaciones anteriores quedarán más claras con algunas ilustraciones. Tomemos el caso de n = 10 líneas. Por lo tanto (n-1)/2 = 4.5 y habrá diseños dialélicos para b = 2, 3 y 4. El dialélico completo sucede cuando 2b = 9. La tabla de cruzas y la matriz correspondiente las presentamos para b = 3, en los diagramas 1, 2, 3 y 4.

Diagrama 4. Formación de la Matriz A

$$A = \begin{bmatrix} 6 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 6 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 6 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 6 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 6 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 6 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 6 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 6 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 6 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 6 \end{bmatrix}$$

$$\sum_{j=1}^{10} a_{1j}^2 = 6^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 = 42$$

$$\sum_{j=1}^{10} a_{1j}\, a_{2j} = 6\times1 + 1\times6 + 1\times1 + 1\times1 + 1\times1 + 1\times1 = 16$$

Sustituyendo valores en la fórmula (13)

$$9\times t = 2\times3\times10 - \frac{42 + 9 \times 16}{4 \times 3} = 60 - 15.5 = 44.5$$

$$\therefore\ t = 4.94$$

Dialélico Completo. Se tiene cuando cada línea se cruza con todas las n-1 restantes y entonces b=(n-1)/2. Las fórmulas que hemos descrito son desde luego, aplicables. Sin embargo, existe una solución inmediata.

La matriz A^{-1} es simétrica con términos (n-1) en la diagonal y 1 e^{n} los demás. Es decir,

$$a_{ii} = n - 1$$
$$a_{ij} = 1,\ i \neq j. \tag{14}$$

La matriz A^{-1} es también simétrica y con elementos

$$a^{ii} = \frac{2n - 3}{2\,(n-1)\,(n-2)} \tag{15}$$
$$a^{ij} = \frac{-1}{2\,(n-1)\,(n-2)},\ i \neq j.$$

El valor de t para este caso es:

$$t = n - 2,$$

como puede comprobarse utilizando la fórmula (13).

Otros diseños parciales. La familia de diseños dialélicos antes descrita no agota las posibilidades. Pueden construirse otros en los que algunas líneas particulares tengan más cruzamientos que otras. Pero aún bajo la restricción de que todas las líneas tengan el mismo número de cruzamientos puede haber otros arreglos. Para ilustrar esta idea consideramos que se tienen n líneas, las que por su origen u otra razón se dividen en dos grupos, n_1 líneas del primer grupo y n_2 líneas del segundo, de tal manera que $n_1 + n_2 = n$. Ahora bien, tenemos todas las cruzas posibles dentro de cada grupo y cruzas parciales entre líneas de los dos grupos. Este caso se presenta en la ilustración siguiente para $n=10$ y $n_1=n_2=5$.

	1	2	3	4	5	6	7	8	9	10
1		x	x	x	x	x	x			
2			x	x	x		x	x		
3				x	x			x	x	
4					x				x	x
5						x				x
6							x	x	x	x
7								x	x	x
8									x	x
9										x
10										

Resumen

Un diseño dialélico está caracterizado por el número de líneas n y por el número b . b = 2,3,... hasta (n-1)/2. El número de cruzas que incluye

el diseño es bn y cada línea se cruza con 2b líneas diferentes. Habrá tantos dialélicos parciales como números enteros hay entre 1 y (n-1)/2, sin incluir a los dos límites 1 y (n-1)/2. El dialélico parcial mínimo ocurre cuando b=2.

Para la construcción del diseño, es decir, definir las cruzas incluidas se utilizan las condiciones (1) o se aplica el método gráfico ilustrado en el Cuadro 1.

Hemos supuesto para simplicidad que las bn cruzas se arreglan en un diseño de bloques al azar. Su análisis de varianza es el siguiente:

Cuadro 3. Análisis de la varianza. Diseño Dialélico n y b.

Fuente de variación	*G. L.*	*S. C.*	*E. C. M.*
Repeticiones	r—1	R	
Aptitud combinatoria general	n—1	L	$\sigma^2 + r\sigma_s^2 + t\sigma_g^2$
Aptitud Combinatoria específica	n(b—1)	S	$\sigma^2 + r\sigma_s^2$
Cruzas	bn—1	C	
Error experimental	(nb—1)(r—1)	E	σ^2
T o t a l :	nbr--1	T	

En el Cuadro 3 las sumas de cuadrados R, C, T, y E se calculan con el método usual. Para el cálculo de L se obtienen primeramente la matriz A, por el método (10) o por el gráfico. En seguida se obtienen las $\hat{Gi}$ con la fórmula (11) y posteriormente L con el auxilio de (12). S se obtiene por diferencia: S = C-L.

Para tener la columna de los valores esperados de los cuadrados medios, E.C.M., el único valor nuevo por calcular es t y éste se obtiene con fórmula (13).

Bibliografía

1. KEMPTHORNE, O Y R. N. CURNOW. "The partial diallel cross". *Biometrics*. (17): 229--250. 1961.
2. GRAYBILL, FRANKLIN A. An introduction to linear statistical models. New York. McGraw Hill. 1961.
3. RAO, C. R. Linear statistical inference and its applications. New York. John Wiley. 1965.