

**ANALISIS TEORICO DE LA EFICIENCIA DEL METODO DE MINIMOS CUADRADOS EN LA SELECCION MASAL**

Jaime Sahagún Castellanos<sup>1</sup>

**RESUMEN**

La heterogeneidad del suelo es reconocida universalmente como uno de los principales factores que enmascaran el valor verdadero de los genotipos cuando se conduce algún tipo de selección. El problema se magnifica cuando para evaluar genotipos sólo se tiene el valor fenotípico de una sola planta por cada genotipo. En este caso, existen varios procedimientos de selección diseñados para controlar al menos una parte de la heterogeneidad del suelo. Uno de tales métodos consiste en practicar la selección con base en valores fenotípicos ajustados substrayendo a los valores fenotípicos observados, la diferencia entre la media fenotípica de las plantas en el sublote correspondiente y la media de ese sublote predicha con base en un modelo de regresión donde las variables independientes son las coordenadas del sublote en el plano bidimensional.

En este estudio se analiza tal método, desde un punto de vista teórico. Los resultados más sobresalientes indican que: 1) El método no controla eficientemente la variación ambiental entre los sublotes y 2) El método le confiere linealidad a los efectos de los sublotes en las coordenadas. Por otra parte, se propone una modificación al método con el propósito de lograr un control completo de la variación entre los sublotes y se discute su eficiencia respecto a otros métodos comparables de selección.

**SUMMARY**

Soil heterogeneity has been universally recognized as one of the main factors that mask the true genotypic value when some type of selection is undertaken. This problem is magnified when the genotypes are evaluated on the basis of the phenotypic value of a single plant. For this situation, there are several selection procedures designed to control, at least partially, the soil heterogeneity. One of such procedures involves the selection on the basis of adjusted phenotypic values. This adjustment consists of subtracting from the actual phenotypic values the difference between the phenotypic mean of the plants in the corresponding block and the predicted mean of this block on the basis of a regression model of the phenotypic values on the coordinates of the bidimensional plane.

In this study such method is analyzed, from a theoretical point of view. Main results indicate that: 1) This method does not control the soil variation among blocks and 2) The method makes linear the block effects on the coordinates. A modification to this method is proposed and its relative efficiency is discussed.

---

<sup>1</sup> Profesor Investigador; Departamento de Fitotecnia; Universidad Autónoma Chapin-  
go, 56230 Chapingo, Edo. de México.

## INTRODUCCION

Un problema que frecuentemente tiene que enfrentar el fitomejorador es la estimación de valores genotípicos. Una estimación eficiente de tales valores generalmente resulta ser laboriosa, costosa y difícil de hacer debido a la acción enmascaradora causada por varios factores, entre los que sobresalen los siguientes: (1) Acción génica no aditiva que pudiera estar presente, por ejemplo, cuando se evalúan líneas con pocas generaciones de autofecundación, o bien, cuando se conduce alguna forma de selección masal; (2) La interacción genético-ambiental, de fuertes implicaciones en la identificación de genotipos para alta adaptabilidad; (3) La heterogeneidad del suelo, cuya presencia es universalmente reconocida y (4) Limitaciones en la cantidad de semilla que representa cada unidad de selección, que con frecuencia impide su evaluación en varias repeticiones, localidades y años.

Este último factor ocurre frecuentemente cuando se evalúan líneas con pocas generaciones de autofecundación. El caso extremo, sin embargo, se presenta cuando los genotipos que requieren de evaluación están representados, cada uno, por una sola planta.

Muchos métodos de mejoramiento (selección masal, método genealógico, selección combinada, la mayoría de los esquemas de mejoramiento de pastos perennes, etc.) involucran la evaluación de genotipos con base en el valor fenotípico de plantas individuales. En estas circunstancias, resulta de capital importancia el control que se pueda ejercer sobre la heterogeneidad del suelo.

Gardner (1961) propone la división del lote de selección en sublotes para que la selección se practique sobre los valores fenotípicos de las plantas individuales dentro de cada sublote. Este método (MC) presenta la ventaja de eliminar la parte de la variación ambiental causada por las diferencias ambientales entre los sublotes. Al controlarse esta componente, habrá una reducción en la varianza fenotípica de las unidades de selección que, en consecuencia, debe redundar en un incremento de la heredabilidad del carácter bajo selección.

En adición a la estratificación, Molina (Márquez, 1971) propone un método (MM) de ajuste para cada valor fenotípico; este método consiste en la substracción de la diferencia entre la media fenotípica del sublote correspondiente y la media fenotípica general. Márquez (1971), como resultado de este ajuste, encuentra que la selección practicada sobre los valores ajustados se realiza como si todas las plantas hubieran gozado de un mismo ambiente. Recientemente, Sahagún

(1983) encuentra que con el método MM, suponiendo despreciable la heterogeneidad dentro de cada sublote, los mejores genotipos de todo el lote de selección son los seleccionados, en tanto que cuando el ajuste no se practica, no hay garantía de que los mejores genotipos de cada sublote sean seleccionados, pudiéndose obtener la máxima eficiencia solamente cuando, por azar, los genotipos superiores se encuentren uniformemente distribuidos en los sublotes. García (1981), trabajando con maíz (*Zea mays* L.), en dos ciclos de selección, obtuvo ganancias mayores cuando la selección se basó en valores ajustados (MM) que al practicar la selección dentro de cada sublote con base en sus valores fenotípicos no ajustados (MC).

Tanto el planteamiento teórico de Sahagún (1983) como el estudio empírico de García (1981) sugieren que el ajuste por efecto de sublotes puede permitir una mejor evaluación de los genotipos. García (1981), en congruencia con esta idea, describe una técnica de selección masal (MG) en la que los valores fenotípicos de cada planta en un sublote son ajustados mediante un procedimiento diferente; a cada valor fenotípico se le resta la diferencia entre la media fenotípica observada de las plantas en el sublote correspondiente y el valor medio predicho mediante la aplicación del principio de mínimos cuadrados a la ecuación de regresión lineal

$$Z_{ij} = \alpha_0 + \alpha_1 X_i + \alpha_2 Y_j + E_{ij} \quad \begin{array}{l} i = 0,1,2, \dots, (p-1) \\ j = 0,1,2, \dots, (q-1) \end{array} \quad (1)$$

donde  $Y_j$  y  $X_i$  representan las coordenadas geográficas del sublote  $ij$ -ésimo;  $Z_{ij}$  es el promedio de los valores fenotípicos de las plantas en el bloque  $ij$ -ésimo;  $\alpha_0$  es la intercepción al origen;  $\alpha_1$  y  $\alpha_2$  son los coeficientes de regresión de las variables  $X$  y  $Y$ , respectivamente, sobre  $Z$ , y  $E_{ij}$  es un término aleatorio de error.

El método MG, por el tipo de ajuste, sugiere la posibilidad de un refinamiento que podría permitir un incremento en su eficiencia. El objetivo del presente estudio consiste en analizar este procedimiento de selección con el fin de determinar su eficiencia en relación a la de otros procedimientos comparables de selección.

#### DESCRIPCION DEL METODO

Varios autores (Trueba, 1971; Flores, 1981) se han referido al empleo de la aplicación de la teoría "mínimo cuadrática" en el problema de la nivelación de tierras. La idea fundamental consiste en hacer la nivelación en forma tal que los cortes de terrenos que se tengan que hacer en las partes elevadas, sean exactamen-

te los necesarios para hacer los rellenos requeridos por las partes más bajas. En términos de selección genotípica, los datos de elevación del terreno que se usan para la nivelación, corresponden a los rendimientos medios de los sublotos del lote de selección, o bien, a la media de cualquier otro carácter para el cual se esté seleccionando.

El método mínimo cuadrático aplicado al modelo (1), proporciona los estimadores de los parámetros  $\alpha_0$ ,  $\alpha_1$  y  $\alpha_2$  o sea  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$  y  $\hat{\alpha}_2$ , respectivamente. Estos estimadores permiten establecer la ecuación de predicción

$$\hat{Z}_{ij} = \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 Y_j \quad (2)$$

que define un plano en el espacio tridimensional que toca al eje Z en el punto  $\hat{\alpha}_0$  y tiene las pendientes  $\hat{\alpha}_1$  y  $\hat{\alpha}_2$  en el sentido de las X's y de las Y's, respectivamente.

El modelo (1) expresado en términos matriciales toma la forma

$$Z = X\alpha + E \quad (3)$$

en donde Z es un vector de orden  $p \times 1$  de la forma

$$Z' = (Z_{00} \ Z_{01} \dots Z_{0p-1} \ Z_{10} \ Z_{11} \dots Z_{1p-1} \dots Z_{q-10} \ Z_{q-11} \dots Z_{q-1p-1});$$

X es una matriz de orden  $p \times 3$ , no singular, de la forma

$$\begin{array}{ccc} J & 0J & C \\ J & J & C \\ J & 2J & C \\ \vdots & \vdots & \vdots \\ J & (p-1)J & C \end{array}$$

donde J es un vector de orden  $q \times 1$  cuyos elementos son la unidad y C es un vector  $q \times 1$  de la forma  $C' = (0, 1, 2, \dots, q-1)$  en tanto que  $\alpha$  es un vector  $3 \times 1$  de la forma  $\alpha' = (\alpha_0, \alpha_1, \alpha_2)$  y E es un vector de errores de orden  $p \times 1$ .

Para la obtención de los estimadores  $\hat{\alpha}_0$ ,  $\hat{\alpha}_1$  y  $\hat{\alpha}_2$  en forma explícita, el método de mínimos cuadrados requiere de la solución del sistema de ecuaciones normales

$$X'X\alpha = X'Z \quad (4)$$

la cual es de la forma

$$\hat{\alpha} = (X'X)^{-1} X'Z \quad (5)$$

Dada la estructura de la matriz  $X$ ,  $X'X$  resulta ser, explícitamente

$$X'X = \begin{bmatrix} pq & q\sum i & p\sum j \\ q\sum i & q\sum i^2 & \sum i\sum j \\ p\sum j & \sum i\sum j & p\sum j^2 \end{bmatrix}$$

en donde las sumas de las  $i$ 's y de las  $j$ 's se refieren a valores de  $i = 0, 1, 2, \dots, p-1$  y  $j = 0, 1, 2, \dots, q-1$ .

Realizando las sumas, la expresión para  $X'X$  queda como

$$X'X = \begin{bmatrix} pq & pq(p-1)/2 & pq(q-1)/2 \\ qp(p-1)/2 & qp(p-1)(2p-1)/6 & pq(p-1)(q-1)/4 \\ pq(q-1)/2 & pq(p-1)(q-1)/4 & pq(q-1)(2q-1)/6 \end{bmatrix}$$

cuyo inverso,  $(X'X)^{-1}$ , es

$$(X'X)^{-1} = \frac{(pqab)^{-1}}{(cd+ab-ad-bc)} \begin{bmatrix} ab(cd-ad) & -ab(d-b) & ab(a-c) \\ -ab(d-b) & b(d-b) & 0 \\ ab(a-c) & 0 & a(c-a) \end{bmatrix}$$

en donde  $a = (p-1)/2$ ,  $b = (q-1)/2$ ,  $c = (2p-1)/3$  y  $d = (2q-1)/3$ .

Usando este resultado, la solución al sistema (4) resulta ser:

$$\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \frac{(pqab)^{-1}}{(cd+ab-ad-bc)} \begin{bmatrix} ab(cd-ab)Z_{..} - ab(d-b) \sum i Z_{i.} + ab(a-c) \sum j Z_{.j} \\ - ab(d-b)Z_{..} + b(d-b) \sum i Z_{i.} \\ ab-(a-c)Z_{..} + a(c-a) \sum j Z_{.j} \end{bmatrix} \quad (6)$$

en donde las expresiones  $Z_{..}$ ,  $Z_{i.}$  y  $Z_{.j}$  indican la suma sobre los subíndices que han sido reemplazadas por los puntos.

Con los estimadores explícitamente expresados en (6), se genera la ecuación de predicción

$$Z_{ij} = \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 Y_j \quad (7)$$

que finalmente, permite realizar el ajuste propuesto para la planta  $k$ -ésima ubicada en el sublote  $ij$ -ésimo ( $MGY_{ijk}$ ) de acuerdo con la expresión

$$MGY_{ijk} = Y_{ijk} - (Z_{ij} - \hat{Z}_{ij}) \quad (8)$$

en donde  $Y_{ijk}$  representa el valor fenotípico observado de la planta  $k$ -ésima del sublote  $ij$ -ésimo.

## DISCUSION

La ecuación de predicción (7) genera un plano en el espacio tridimensional en donde los puntos  $\hat{Z}_{ij}$  (para valores  $X_i = 0, 1, 2, \dots, p-1$ ;  $Y_j = 0, 1, 2, \dots, q-1$ ) son las medias predichas de la variable de interés, correspondientes a los diferentes sublotes. Con el ajuste indicado por la expresión (8), es evidente que todas las plantas de un mismo sublote son ajustadas por una misma cantidad. Suponiendo que las diferencias entre las medias fenotípicas reales de los sublotes son atribuibles a efectos ambientales exclusivamente, el efecto del ajuste consistiría en "mover" a cada media la cantidad en que su media observada exceda a la media predicha correspondiente. Si esta diferencia es positiva, los valores fenotípicos de las plantas en el sublote experimentarían un ajuste que consiste en reducir una cantidad igual a la diferencia observada en ese sublote; en tanto que si el "exceso" es negativo, los valores fenotípicos de las plantas en ese sublote sufrirían un incremento igual al valor absoluto de tal exceso.

Bajo estas circunstancias resulta obvio que si la selección se practicara dentro de cada sublote, entonces este método de ajuste proporcionaría los mismos resultados que los que se obtendrían mediante la técnica de selección propuesta por Gardner (1961). Esta última, sin embargo, sería más conveniente en virtud de no requerir los cálculos necesarios para hacer las estimaciones de los parámetros  $\alpha_0$ ,  $\alpha_1$  y  $\alpha_2$  ni de requerir el ajuste. Por otra parte, el método de Gardner (MC) tampoco impone la restricción de que el lote de selección sea rectangular. En cambio, si la selección se practicara sobre los valores fenotípicos ajustados únicamente, independientemente de los sublotes en que se encuentren las plantas, como fue propuesto originalmente, entonces los efectos ambientales podrían confundirse con los genotípicos: las plantas localizadas en los sublotes cuyos valores medios predichos se localizan en las partes superiores del plano, tendrían mayor probabilidad, por ese solo hecho, de ser seleccionadas que las que estuvieran en las partes bajas. La eficiencia sería menor a medida que las estimaciones de los coeficientes de regresión tomaran valores absolutos mayores.

Una forma de corregir este problema sería mediante un reajuste (MR) consistente en "mover" el plano de tal manera que las pendientes de éste sean cero. En términos matemáticos, el valor fenotípico de la k-ésima planta así reajustada ( $MRY_{ijk}$ ) tendría la expresión

$$MRY_{ijk} = Y_{ijk} - (Z_{ij} - \hat{Z}_{ij}) - (\hat{\alpha}_1 X_i + \hat{\alpha}_2 Y_j) \quad (9)$$

que en forma más breve es

$$MRY_{ijk} = Y_{ijk} - Z_{ij} + \hat{\alpha}_0 \quad (10)$$

dado que

$$\hat{Z}_{ij} = \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 Y_j$$

Desde el punto de vista de la selección, seleccionar con base en los valores fenotípicos ajustados descritos en (10) conduciría a seleccionar los mismos genotipos que se seleccionarían si la selección se basara en los valores fenotípicos ajustados

$$MSY_{ijk} = Y_{ijk} - Z_{ij} \quad (11)$$

Esto es así porque al ser  $\hat{\alpha}_0$  un valor constante en todo valor  $MRY_{ijk}$ , las diferencias del tipo

$$MRY_{ijk} - MRY_{i'j'k'}$$

para toda  $i, j, k, i', j',$  y  $k'$ , son iguales a las diferencias correspondientes de la forma

$$MSY_{ijk} - MSY_{i'j'k'}$$

Evidentemente, el fitomejorador desearía aplicar el ajuste descrito por la expresión (11) ya que éste no requiere de ningún cálculo en términos de análisis de regresión.

Por lo que respecta a la eficiencia del método de selección basado en los valores fenotípicos ajustados según la expresión (11), es de notarse que este método es equivalente al MM, pues si bien éste basa su ajuste en valores fenotípicos ajustados ( $MMY_{ijk's}$ ) dados por la expresión

$$MMY_{ijk} = Y_{ijk} - (Z_{ij} - \bar{Z}),$$

en donde  $\bar{Z}$  es la media de todos los valores fenotípicos en el lote de selección, es claro que este procedimiento conduce a la selección de los mismos genotipos que el método basado en los valores fenotípicos ajustados según la expresión (11), puesto que al diferir ambos tipos de valores ajustados en una constante ( $\bar{Z}$ ) el orden de las magnitudes de ambos tipos de valores fenotípicos ajustados es el mismo.

El ejemplo siguiente, con datos hipotéticos, se presenta con el fin de ilustrar

trar los conceptos anteriores. Supóngase que se tiene un lote de selección cuyos sublotos tienen las medias y coordenadas descritas en el Cuadro 1. Para este caso  $p=5$ ,  $q=5$ ,  $a=2$ ,  $b=2$ ,  $c=3$  y  $d=3$ , de donde, por (6)

$$\hat{\alpha}_1 = \frac{100^{-1}}{(9+4-6-6)} \{-4(200) + 2(1 \times 45 + 2 \times 40 + 3 \times 35 + 4 \times 30)\}$$

$$= \frac{1}{100} \{-800 + 2(350)\}$$

$$= -1$$

$$\hat{\alpha}_2 = \frac{1}{100} \{-4(200) + 2(1 \times 30 + 2 \times 40 + 3 \times 50 + 4 \times 60)\}$$

$$= \frac{1}{100} \{-800 + 2(500)\}$$

$$= 2$$

$$\hat{\alpha}_0 = \frac{1}{100} \{4(5)(200) - 4(350) + 4(-1)(500)\}$$

$$= \frac{1}{100} \{4000 - 1400 - 2000\}$$

$$= 6$$

Cuadro 1. Medias y coordenadas de los sublotos que forman un lote de selección hipotético.

C o o r d e n a d a s						
$x_i$	$y_j$					Total
	0	1	2	3	4	
0	6	8	10	12	14	50
1	5	7	9	11	13	45
2	4	6	8	10	12	40
3	3	5	7	9	11	35
4	2	4	6	8	10	30
Total	20	30	40	50	60	200

Para ilustrar las diferencias entre el método MG (8) y el método MR (10) considérense los valores que tendrían que substraerse de cada valor fenotípico observado según cada procedimiento. Los valores fenotípicos correspondientes a las plantas del sublote  $ij$ -ésimo se ajustarían de acuerdo con la expresión

$$\begin{aligned}
 MGY_{ijk} &= Y_{ijk} - (Z_{ij} - \hat{Z}_{ij}) \\
 &= Y_{ijk} - (Z_{ij} - \hat{\alpha}_0 - \hat{\alpha}_1 X_i - \hat{\alpha}_2 Y_j) \\
 &= Y_{ijk} - (Z_{ij} - 6 + X_i - 2Y_j)
 \end{aligned}$$

en tanto que bajo MS, el ajuste sería

$$MRY_{ijk} = Y_{ijk} - Z_{ij} + \hat{\alpha}_0 \quad i, j = 0, 1, 2, 3, 4$$

Las cantidades que habría que sustraer de cada uno de los valores fenotípicos observados al aplicar los métodos MG y MR en cada uno de los sublotes se presentan en el Cuadro 2.

Cuadro 2. Valores que tendrían que restarse de los valores fenotípicos observados del Cuadro 1 al aplicar los métodos MG y MR en un caso hipotético de selección.

$X_i$	$Y_j$					$X_i$	$Y_j$				
	0	1	2	3	4		0	1	2	3	4
	MG						MR				
0	0	0	0	0	0	0	6	8	10	12	14
1	0	0	0	0	0	1	5	7	9	11	13
2	0	0	0	0	0	2	4	6	8	10	12
3	0	0	0	0	0	3	3	5	7	9	11
4	0	0	0	0	0	4	2	4	6	8	10

Estas cantidades indican claramente que al aplicar el método MG en realidad no hubo ajuste por efecto de las diferencias ambientales entre los sublotes, puesto que el "ajuste" consistió en sustraer de cada valor fenotípico una cantidad constante (el cero) ya que en este ejemplo los valores reales son iguales a los predichos, pues no hay desviaciones de regresión. Por supuesto, al aplicar el método MR, los valores ajustados obtenidos, de acuerdo con la expresión (10), resultan ser los valores que se obtendrían si las plantas se hubieran desarrollado en sublotes que no difieren en sus efectos. Este aspecto implicaría, por lo tanto, un control perfecto de la variación ambiental entre los sublotes. De esta manera, si

además no hubiera variación ambiental intrasublotte, esta metodología de selección permitiría la detección de los genotipos más sobresalientes del lote de selección de acuerdo con la presión de selección establecida.

En una situación real, obviamente, lo común sería encontrar discrepancias entre los valores observados y los valores predichos. Pero aun en este caso, el ajuste con el método MG consistiría en adicionar a los valores fenotípicos las cantidades necesarias para que las medias de los sublotes correspondientes queden en el plano formado por los puntos  $\hat{Z}_{ij}$ , y no en un ajuste que controle la variación ambiental entre los sublotes. Este ajuste se logra sólo cuando se da el plano con pendiente cero obtenido con el método MR.

Debe hacerse énfasis en que los métodos aquí discutidos sólo están planeados para ejercer algún control sobre la variación ambiental entre los sublotes. Si los sublotes fueran aceptablemente homogéneos dentro de ellos pero hubiera considerable variación entre los mismos, los ajustes deberían producir incrementos en eficiencia en relación a la utilización de los valores fenotípicos no ajustados como un criterio de selección. El problema consistiría entonces en el diseño de una metodología que permitiera determinar el grado de heterogeneidad dentro de los sublotes o bien que permitiera la posibilidad de ejercer algún control sobre tal variación. Si todos los efectos debidos a la heterogeneidad del suelo pudieran ser controlados, entonces, para un ambiente, la evaluación de los genotipos con base en el valor fenotípico de una sola planta sería de una eficiencia muy alta. Desafortunadamente, no existe procedimiento alguno que permita tal grado de control. No obstante, algunos otros procedimientos deberían ser investigados mediante el uso de datos experimentales. Un ejemplo de estos procedimientos es el propuesto por Méndez y Rivera (1976) el cual fue descrito en términos teóricos solamente.

#### BIBLIOGRAFIA

- Flores A., R. 1981. Evaluación de diferentes métodos de nivelación de tierras y su programación en el lenguaje Fortran. Tesis de Licenciatura. Universidad Autónoma Chapingo. 198 p.
- García G., J.C. 1981. Comparación de tres técnicas de selección masal en un compuesto precoz de maíz. Tesis de Maestría en Ciencias. Colegio de Postgraduados. Chapingo, México.
- Gardner, C.O. 1961. An evaluation of the effects of mass selection and seed irradiation with thermal neutrons on yield of corn. Crop Sci, 1:241-245.

- Márquez S., F. 1971. Interpretación a la fórmula de ajuste del rendimiento individual en la selección masal. *Fitotecnia* 2:1-2.
- Méndez R., I. y G. Rivera H. 1976. Métodos alternativos para comparación de un número grande de variedades, usando funciones de tendencia. *Agrociencia* 26:153-165.
- Sahagún C., J. 1983. Comparación teórica de las eficiencias de dos técnicas de selección masal. *Fitotecnia* 5:43-57.
- Trueba C., S. 1971. Coeficientes de Trueba para el Cálculo de Nivelación de Tierras Agrícolas. Editorial CECSA. México pp 18-38.