SEMINARIO DE ECONOMIA _.

(26)

ANALISIS DE DATOS CORRELACIONADOS
RECUPERACION DE LA ESTRUCTURA DE CORRELACION INTRABLOQUE.

María Gabriela Benedicto[1]

RESUMEN

El diseño experimental clásico se basa en los conceptos de aleatorización, bloqueo y replicación. La aleatorización conduce a neutralizar los efectos de la correlación espacial y proporciona tests válidos para la hipótesis de igualdad de tratamientos. Sin embargo, más recientemente se han hecho intentos de recuperar la información de esa estructura de correlación para mejorar la eficiencia de los estimadores.

En este trabajo se presenta una metodología alternativa para la estimación de la función de variograma (varianza de la diferencia de tratamientos que están a distancia h), basada en el postbloqueo de los ensayos. Esta técnica consiste en imponer al ensayo ya realizado, bloques de distinto tamaño y estimar los cuadrados medios de error como si éste fuera el diseño original. Se probará que, bajo ciertos supuestos, el CME es efectivamente un estimador insesgado de la función de variograma y se lo comparará con el estimador clásico calculado por mínimos cuadrados ordinarios.

Finalmente se presenta una aplicación de esta metodología a una serie de ensayos reales y se propone un tamaño de bloque adecuado de acuerdo al comportamiento espacial de los tratamientos.

I- INTRODUCCION

Uno de los objetivos de los organismos de investigación agropecuarios, como por ejemplo el INTA, es la evaluación de nuevas líneas y variedades o la comparación de distintos tratamientos. El diseño usualmente utilizado, aunque no por ello el óptimo, es el de Bloques Completos Aleatorizados.

El diseño experimental clásico se basa en los conceptos de aleatorización, bloqueo y replicación. La aleatorización conduce a neutralizar los efectos de la correlación espacial y proporciona tests válidos para la hipótesis de igualdad de tratamientos. Sin embargo, se puede ver que las parcelas vecinas tienden a ser más parecidas que las más distantes. Si se encuentra que las parcelas cercanas están muy correlacionadas, el incorporar esta información en un análisis de tipo "vecino" o "espacial" puede llevar a obtener estimadores más eficientes de las diferencias varietales que posiblemente los concluídos de un análisis estándar de bloques.

Así, un aspecto importante en la etapa de análisis de los resultados es la estimación del variograma ($2\gamma(x)$), es decir, de la varianza de la diferencia de tratamientos en parcelas que distan x.

Las estimaciones de los variogramas son de tipo noparamétrico y están basadas en los residuales del modelo planteado para el análisis del experimento.

Dos estimadores conocidos son:

---

[1]-Departamento de Economía - Universidad de San Andrés

(i) estimador por el método de los momentos (Matheron, 1963)

$$2\hat{\gamma}(x) = \sum_{N(x)} (R(s_i)-R(s_j))^2 \;/\; |N(x)|$$

(ii) estimador robusto (Cressie y Hawkins, 1980)

$$2\overline{\gamma}(x) = \{\sum_{N(x)} |R(s(i)-R(s(j)|^{1/2}/|N(x)|\}^4/\{0.457+0.494/|N(x)|\}$$

donde $N(x)=\{(s_i,s_j): s_i - s_j = x\}$, y $|N(x)|$ es el número de pares en $N(x)$.

Una forma alternativa para la estimación de los variogramas es obtenerlos a partir de los CME calculados para distintos tamaños de bloques ya que se puede probar que estos valores están íntimamente relacionados. Esta técnica recibe el nombre de Postbloqueo y en ella estarán basadas las estimaciones presentadas en este trabajo.

La relación que liga los CME y los variogramas es la siguiente:

$$E(CME(k)) = \frac{2\sum_{x=1}^{k-1}(k-x)\,E(\varepsilon_{ij}-\varepsilon_{i,j+k})^2}{k(k-1)} \tag{1}$$

(Patterson y Hunter ,1983; Ainsley ,1985; Grondona y Cressie, 1991). Esta esperanza se toma sobre la aleatorización y sobre el modelo de los ϵ_j, es decir, se calcula

$$E\,(CME(k)) = E_A\,(\,E_M\,(\,CME\,/\,Aleatorización)).$$

Vale aclarar que desde el punto de vista de la aleatorización, el postbloqueo origina tests válidos, esto es, E(CMTrat)=E(CME) bajo la hipótesis nula de que no hay efectos de tratamientos, (Ainsley, Paterson y Patterson, 1987).

Se compararán también las estimaciones derivadas a partir del Método de los Momentos, que es la que se usa habitualmente, y la que se concluye de la aplicación de la técnica de Postbloqueo.

II- ESTIMACION DE LOS VARIOGRAMAS

Planteemos el modelo

$$y_{jk} = \mu + \beta_j + \alpha_k + \omega_{jk} + \varepsilon_{jk} \tag{2}$$

donde μ = media general , β_j= efecto de bloques (j:1,...,b) , α_k = efecto de tratamientos (k:1,...,t), ε_{jk}= ruido blanco y ω_{jk} es el error producido al introducir la aleatorización, con $E(\omega_{jk})=0$, $Var(\omega_{jk})=\sigma^2_{\omega j}$ y $Cov(\omega_{jk},\omega_{jk'})=-\sigma^2_{\omega j}/(t-1)$.

Este modelo puede derivarse a partir de un enfoque espacial, (Grondona y Cressie ,1991), y suponiendo la aleatorización de los tratamientos en un diseño en bloques completos.

Calculando la tabla de análisis de varianza de este modelo, y tomando esperanza al Cuadrado Medio de Residuales obtenemos, cuando no hay error de medición,

$$E(CME) = \frac{1}{r(t-1)} \sum_{i=1}^{t} \sum_{j=1}^{r} (\mu_{ij}-\mu_{.j})^2 + \frac{1}{2t(t-1)r} \sum_{j=1}^{r} \sum_{i=1}^{t} \sum_{l=1}^{t} 2\gamma_w(x_{il,j}) \quad (3)$$

donde $h_{il,j}$ representa la distancia entre la i-ésima y l-ésima unidades dentro del bloque j. Por lo tanto, aún cuando se considere error de medición nulo, el CME es un estimador sesgado del semivariograma promedio a menos que se hagan algunos supuestos sobre los efectosde tendencia $\{\mu_{ij}\}$,por ejemplo, $\mu_{ij}= \mu_{.j}$ ó $\mu_{ij}= \mu_{..}$. Este es el supuesto que se hace en este enfoque. Usando postbloqueo con tamaño de bloques variable y teniendo en cuenta esta relación, se pueden derivar estimadores de la función de variograma. (Ainsley, Patterson y Patterson,1987; Patterson y Hunter, 1983).

Veamos ahora qué ocurre con la esperanza del estimador derivado por Mínimos Cuadrados Ordinarios.

Partiendo del modelo anterior, tenemos que la estimación por MCO del variograma es

$$2\hat{\gamma}(x) = \frac{1}{b(t-x)} \sum_{j=1}^{b} \sum_{i=1}^{t-x} (e_{i+x\,j}-e_{ij})^2 \quad (4)$$

donde los e_{ij} son los residuales de la estimación por MCO

Condicionando nuevamente la esperanza del estimador a la aleatorización,obtenemos

$$E_M\{2\hat{\gamma}(x)/A\}=2\gamma(x) + \frac{1-2b}{b^2} 2\gamma(x) + \frac{1}{b^2} \sum_{k=1,\, k\neq j}^{b} Var(y_{I(i+x,j,k),k}-y_{I(i,j,k),k}) \quad (5)$$

donde $I(i,j,k) = \{i' : \tau(i',k)=\tau(i,j)\}$,
y finalmente se llega a que

$$E\{2\hat{\gamma}(x)\} = 2\gamma(x) + \frac{1-2b}{b^2} 2\gamma(x) + \frac{b-1}{b^2 t(t-1)} \sum_{x'=1}^{t-1} (t-x') 2\gamma(x') \quad (6)$$

De esta ecuación concluimos que el estimador del variograma derivado por MCO no es insesgado, sino que su valor va a depender del número de bloques y de tratamientos.

De lo demostrado hasta aquí se podría concluir que el estimador del variograma obtenido al aplicar la técnica de postbloqueo es más exacto, pero no por ello podemos decir que sea más eficiente ya que se puede ver que las estimaciones presentan una gran variabilidad, conduciendo incluso, algunas veces, a valores negativos.

III- APLICACION DE LA TECNICA

Se trabajó con datos correspondientes a un ensayo de uniformidad realizado por Mercer y Hall en la Estación Experimental de Rothamsted en 1910. Estos estaban originalmente dispuestos en un látice de 20x25 parcelas y cubrían aproximadamente un área total de un acre. A estos datos se les impusieron artificialmente seis diseños en bloques completos aleatorizados, cada uno con 20 tratamientos y 4 bloques, Grondona (1989).

En el trabajo citado, se estimó el variograma de los datos originales y se le ajustó una función de tipo exponencial ,Figura 1.

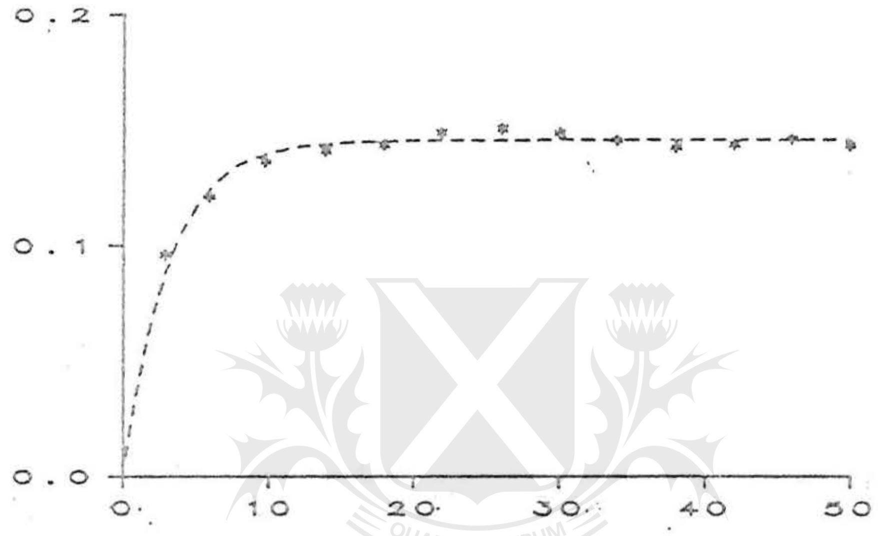

Figura 1: Ajuste por Mínimos Cuadrados de la función de variograma del ensayo de uniformidad de Mercer y Hall

Se podría esperar que una vez impuestos los tratamientos, el variograma se siga comportando aproximadamente de la misma manera, esto es, que por ser un proceso estacionario, exista un punto donde el variograma se estabilice. Si esto ocurre, es decir, si existe un valor de h_0 a partir del cual $\gamma(x)$ es constante, lo mismo ocurre con el CME(k).

Postbloqueando el diseño y estimando los Cuadrados Medios de Error para cada tamaño de bloque, podemos ver que esto no ocurre sino que crecen en forma lineal, Figura 2. Al estimar los variogramas según la ecuación 1, obtenemos la relación presentada en la Figura 3. Esta relación tampoco concuerda con la planteada originalmente en el ensayo de uniformidad, como era de esperar al observar el comportamiento de los CME.

Todo esto nos está indicando que las estimaciones de los variogramas son sesgadas: existe tendencia en los datos. Una forma de solucionar este problema y poder modelar la correlación, es proponer un modelo ARIMA que libere al ajuste de esta tendencia. Gleeson y Cullis (1987) proponen métodos basados en primera y segunda diferenciaciones de los datos. Estos métodos son comparables al uso de diferencias en series de tiempo para lograr estacionalidad.

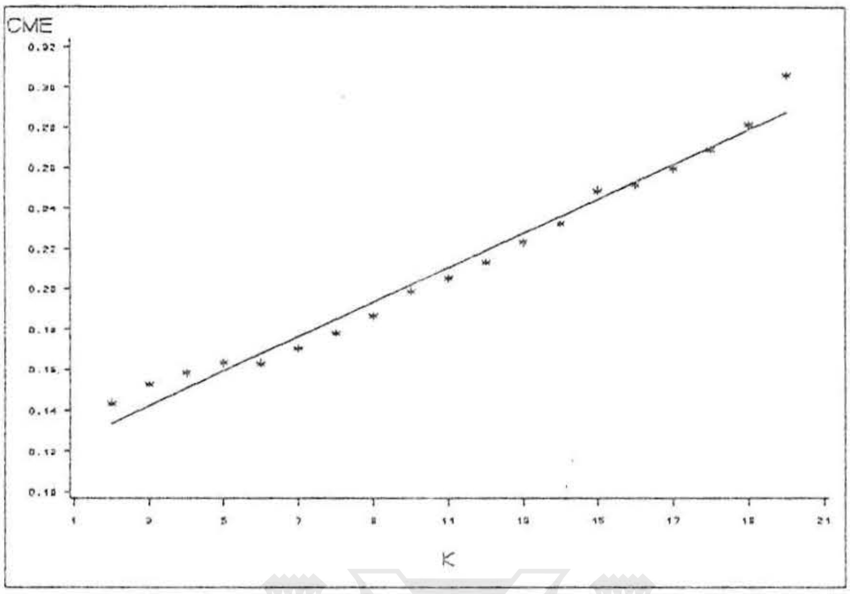

Figura 2: Ajuste de los valores de los Cuadrados Medios de Error obtenidos por la técnica del postbloqueo, en función de los tamaños de bloques.

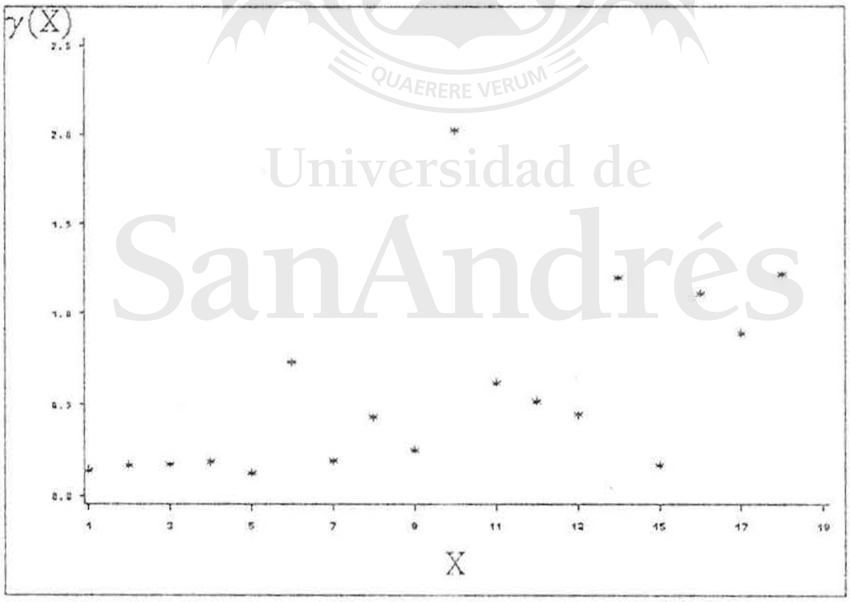

Figura 3 : Variogramas obtenidos a partir de los CME, promediados a través de los seis ensayos.

Si miramos los gráficos detenidamente, podemos ver que hasta el tamaño de bloque 6, ambas funciones se pueden aproximar por una curva de tipo exponencial. A partir de ahí, los datos comienzan a perder estabilidad. Esto podría sugerir que el tamaño óptimo de bloque para este experimento tendría que ser no mayor a k=6 para obtener un buen estimador de la función de variograma y así mejorar la eficiencia del análisis. Esto significa, decidirnos por un diseño en Bloques Incompletos, que evitaría la introducción de nuevas fuentes de variabilidad producidas por trabajar con bloques demasiado grandes.

Esta conclusión coincide con la propuesta de Patterson y Hunter (1983), que sugiere que para la comparación de t variedades, con $t \leq 64$, la mejor elección del tamaño de bloques, k, es tomar $k \approx t^{1/2}$.

III- DISCUSION

La técnica de postbloqueo es una metodología útil para estimar las esperanzas de los CME en función del tamaño de bloque y por lo tanto, de la función de variograma. Estas estimaciones son insesgadas cuando no existe tendencia dentro de los bloques.

El rescatar esta estructura de correlación de las parcelas, permite mejorar las estimaciones en la etapa del análisis del experimento y evaluar la bondad del diseño utilizado. Esto ayudará a elegir el diseño óptimo en experimentos similares futuros.

IV- REFERENCIAS

AINSLEY, A. E. (1985). Unpublished Ph. D. tesis, University of Edinburgh.

AINSLEY, A.E., PATERSON, L.J. Y PATTERSON, H.D. (1987). A method for predicting the efficiency of incomplete-block trials. Biometrics, 43; 55-59.

CRESSIE, N. y HAWKINS, D.M. (1980). Robust estimation of the variogram:I. Journal of the International Association for Mathematical Geology, 12, 115-125.

GLEESON, A.C. y CULLIS, B.R. (1987). Residual Maximun Likelihood (RML) estimation of a neighbour model for field experiments. Biometrics, 43; 277-288.

GRONDONA, M.O. (1989). Estimation and design with correlated observations.Ph.D. Thesis. Iowa State University.

GRONDONA, M.O. y CRESSIE, N. (1991). Using spatial considerations in the analysis of experiments. Technometrics, 33, Nro. 4.

MATHERON, G. (1963). Principles of geostatistics. Economic Geology, 58,1246-1266.

PATTERSON, H.D. y HUNTER, E.A. (1983). The efficiency of incomplete block designs in National List and Recommended Variety Trials. Journal of Agricultural Sience, 101; 427-433.

PATTERSON, H.D., WILLIAMS, E.R. y HUNTER, E.A. (1978). Block design for variety trials. Journal of the Agricultural Science, Cambridge, 90; 395-400.