

Una aplicación del test de solapamiento en un modelo de clasificación de dos vías reparametrizado

An application of the overlapping test in a two-way classification reparameterized model

A.E. Darghan Contreras¹, P. Sinha Surendra² y A. Goitia Acosta²

¹ Universidad Nacional Experimental del Táchira, Departamento de Agronomía. P.O. BOX 436. San Cristóbal, Táchira, Venezuela.

² Universidad de los Andes. Instituto de Estadística Aplicada y Computación. Mérida, Venezuela.

Resumen

En este artículo se presenta una aplicación del score test de Rao para el coeficiente de solapamiento en el modelo de Draper y Guttman utilizando un modelo de clasificación de dos vías y la imposición de restricciones sobre los parámetros para completar el rango de la matriz de diseño. El test score de Rao en este artículo involucra la teoría asociada a los operadores de proyección perpendicular y usa los estimadores máximo-verosímiles restringidos por funciones no estimables. El test puede aplicarse con diferentes patrones de vecindad siempre y cuando se considere al vecino más cercano como la unidad experimental directamente afectada por los tratamientos. El método es de fácil implementación y puede aplicarse en el campo de la agronomía o en temas afines con otras áreas, tal como en el manejo de pastos y forrajes en la producción animal y veterinaria, pues la naturaleza asintótica del test está en concordancia con el gran número de unidades experimentales presentes comúnmente en esta área.

Palabras clave: Solapamiento, Operador de proyección perpendicular, Score Test, Funciones No Estimables, Modelo Reparametrizado.

Abstract

In this article, an application of the score test of Rao is presented for the overlap coefficient in the Draper and Guttman's model using a two-way classification model and the imposition of restrictions on the parameters to complete the rank of the design matrix. The score test of Rao in this article involves

the theory associated with the perpendicular projection operators and uses the maximum likelihood estimators restricted by non estimable functions. The test can be applied with different neighborhood patterns whenever the closest neighbor can be considered as the experimental unit directly affected by the treatments. The method is simple to implement and can be applied in the field of Agronomy or in similar studies in other areas such as pasture and forage management in animal and veterinary production since the asymptotic nature of the test is in agreement with the large number of experimental units commonly present in this area.

Key words: Overlapping, Perpendicular projection operator, Score test, Non estimable functions, Reparameterized model.

Introducción

En la investigación agropecuaria es común encontrar que en ciertos experimentos la aplicación de un tratamiento sobre una unidad experimental algunas veces se dispersa a unidades experimentales adyacentes pudiendo afectar su respuesta, es decir que podría presentarse un solapamiento mutuo en el uso de los tratamientos. Por ejemplo, en ensayos de variedades, el efecto de la vecindad puede ser atribuido a diferencias morfológicas intrínsecas entre las plantas de diferentes variedades o a diferencias en la fecha de germinación inherente a las semillas o inducidas por el ambiente. Los tratamientos aplicados a los cultivos tales como planes de fertilización, riego, pesticidas, uso de controladores biológicos, agentes inoculantes, entre otros, pudieran dispersarse a las parcelas vecinas con lo cual puede afectarse la variable respuesta que está siendo medida en un momento dado. Ejemplos de estas situaciones pueden leerse en Bhalli *et al.* (1964) y en Hide y Read (1990).

El fenómeno de solapamiento ha sido modelado por varios investigado-

Introduction

In an agroindustrial research is common finding that in some experiments the application of a treatment on an experimental unit sometimes scatter to adjacent experimental units which might affect the answers, that is, there may be a mutual overlapping in the usage of treatments. For example, in essays of varieties, the neighborhood effect might be attributing to intrinsic morphological differences between plants of different varieties, or in differences in the germination date inherent to the seeds or induced by the environment. The treatments applied to the crops such as fertilization plans, irrigation, pesticides, use of biological controllers, inoculants agents, among others, may disperse to the neighbor plots, thus affecting the response variable, which is measured on a specific moment. Examples of these situations can be read in Bhalli *et al* (1964) and Hide and Read (1990).

The overlap phenomenon has been modeled by different authors, for example, Pearce (1957) considered a model where each treatment has a

res, por ejemplo, Pearce (1957) consideró un modelo donde cada tratamiento tiene un efecto directo en la parcela en la cual es aplicado y un efecto de solapamiento sobre las unidades experimentales vecinas. Draper y Guttman (1980) estudiaron el solapamiento en superficies de respuesta, específicamente cuando se considera al vecino más próximo (horizontal, vertical o diagonal) como la fuente del traslapo, además, discutieron algunos métodos de prueba aproximados para el coeficiente de solapamiento y usaron el modelo no lineal

$$Y = X\beta + \alpha WX\beta + \varepsilon, \quad (1)$$

donde Y es un vector aleatorio de dimensión $n \times 1$ que denota la respuesta de n unidades experimentales, X es una matriz de diseño conocida de dimensión $n \times p$, β es un vector de parámetros desconocidos de dimensión $p \times 1$ que consiste de los efectos a considerar en el experimento, que en el caso de un modelo de clasificación de dos vías constará del efecto de los tratamientos y el efecto de los bloques, α representa el efecto desconocido del solapamiento, $W = (w_{ijr})$ es una matriz de pesos conocida de dimensión $n \times n$ donde w_{ijr} denota el efecto de la unidad j sobre la unidad i en la réplica r ; con la condición de que $r; \sum_{j=1}^n w_{ijr} = 1 \forall i, r; w_{ijr} = 0$ y $w_{ijr} \geq 0 \forall i, j, r$. Finalmente, para el vector de errores ε se asumió distribución normal e independencia con vector de medias cero y varianza $\sigma^2 I$, donde I es una matriz identidad de dimensión $n \times n$. (Shukla y Subrahmanyam, 1999).

Recientemente, Darghan *et al.*, (2009) propuso el test de solapamiento en modelos de superficies de respuesta

direct effect on the plot where it is applied, and overlapping effect on the neighbors' experimental units. Draper and Guttman (1989) studied the overlapping in response's surfaces, specifically when the nearest neighbor is considered (horizontal, vertical or diagonal) as the overlap's source; also, they have discussed some approximate tests' methods for the overlapping coefficient and used the non lineal method.

$$Y = X\beta + \alpha WX\beta + \varepsilon, \quad (1)$$

Where Y is a random dimension vector $n \times 1$ which implies the response of n experimental units, X is a known design matrix of dimensions $n \times p$, β is an unknown parameter vector of dimension $p \times 1$ which consists on the effects to be considered on the experiment, and on the case of a two-way classification model will have the treatments' effect and the blocks' effect, α represents the unknown overlapping effect, $W = (w_{ijr})$ is a weight known matrix of dimensions $n \times n$ where w_{ijr} indicates the unit's effect j on the unit i and on the replication r ; with the condition that $r; \sum_{j=1}^n w_{ijr} = 1 \forall i, r; w_{ijr} = 0$ and $w_{ijr} \geq 0 \forall i, j, r$.

Finally, for the mistakes' vector "e", it was assumed a normal distribution and independence with vector of zero means and variance " $\sigma^2 I$ ", where I is an identity matrix of dimension $n \times n$ (Shukla and Subrahmanyam, 1999).

Recently, Darghan *et al.*, (2009) proposed the overlapping test in surfaces response's model of first and second order, and Sinha *et al.* (2009) proposed the overlapping test in two-way classification models restricted by

de primer y segundo orden y Sinha *et al.*, (2009) propuso el test de solapamiento en modelos de clasificación de dos vías restringido por una función no estimable basándose en la prueba fundamental de hipótesis desarrollada por Rao (1948) y conocida actualmente como el Test Score de Rao. El test se desarrolló sobre el modelo reparametrizado

$$Y = X_*\gamma + \alpha WX_*\gamma + \varepsilon, \quad (2)$$

donde $X_* = XZ$, es de rango completo por columnas a diferencia de X , $\beta = Z\gamma$ y $c(Z) = c(A)^\perp$ (la última expresión significa que el espacio de columnas de la matriz Z es igual al perpendicular del espacio de columnas de la matriz A , lo que a su vez implica que $A'\beta = 0$).

La reparametrización anterior permite completar el rango de X y a su vez la obtención de estimadores únicos máximo-verosímiles a través del uso de funciones no estimables donde $c(X_*) = c(X)$, es decir, $A'\beta = 0$ involucra solamente el uso de condiciones laterales que no afectan al modelo (1), (Christensen, 2002).

En este artículo se mostrará mediante una aplicación agronómica el uso del Test de Sinha para solapamiento haciendo uso del método de estimación máximo-verosímil con el modelo reparametrizado. En la aplicación del test se usará como ejemplo ilustrativo la misma matriz de datos proporcionada por el autor pero para un patrón de vecinos más cercanos distinto al expuesto en su artículo original, además, se mostrarán con cierto detalle los pasos para la reparametrización del modelo (1). El ejemplo utilizado condujo al rechazo de

a non estimable function based on the hypothesis fundamental test developed by Rao (1948) and currently known as Rao Test Score. The test was developed on the reparameterized model.

$$Y = X_*\gamma + \alpha WX_*\gamma + \varepsilon, \quad (2)$$

Where $X_* = XZ$, is a complete rank by columns different from X , $\beta = Z\gamma$ y $c(Z) = c(A)^\perp$ (the last expression means that the column's space of the matrix Z is equal to the perpendicular space of column of the matrix A , which at the same time implies that $A'\beta = 0$).

The latter reparameterization allows completing the rank X and at the same time the obtaining of maximum-reliability estimators using non estimable functions where $c(X_*) = c(X)$, that is, $A'\beta = 0$ it only involves the usage of lateral conditions that do not affect the model (1), (Christensen, 2002).

In this research, will be shown through an agronomic application the usage of the Sinha test for overlapping, using the maximum-reliability estimation method with the reparameterized model. On the application of the test, will be used as illustrative example the same data matrix provided by the author but for a nearest neighbor pattern different than the exposed on the original article, also, will be shown with some details the steps for the reparameterization of the model (1). The example used carried to the reject of the null hypothesis associated to the absence of overlapping between treatments, thus, a common variance analysis in the tow-way classification model will be questionable.

la hipótesis nula asociada a la ausencia de solapamiento entre tratamientos, por lo cual, un análisis de varianza usual en el modelo de clasificación de dos vías sería cuestionable.

Metodología

Estimación de Parámetros y test de solapamiento de Sinha

Sinha *et al.* (2009) desarrollaron un test de solapamiento usando el método de construcción de pruebas de hipótesis desarrollado por Rao (1948). Este test usa al modelo (2) y el hecho de que en este modelo se tiene también que $\sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, de este modo, la función logaritmo de la verosimilitud viene dada por

$$l(y; \alpha, \gamma, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|\mathbf{y} - \alpha \mathbf{X} \gamma\|^2, \quad (3)$$

donde $Q = I + \alpha W$.

Bajo $H_0: \alpha = 0$, la función logaritmo de la verosimilitud para la variable aleatoria rinde los siguientes estimadores:

$$\tilde{\gamma} = (\mathbf{X}_*^t \mathbf{X}_*)^{-1} \mathbf{X}_*^t \mathbf{y} \quad (4)$$

y

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}_* \tilde{\gamma})^t (\mathbf{y} - \mathbf{X}_* \tilde{\gamma}), \quad (5)$$

donde \mathbf{X}_* , es la matriz de diseño del modelo reparametrizado de modo que se pueda garantizarse la existencia de $(\mathbf{X}_*^t \mathbf{X}_*)^{-1}$.

Rao (1948) introdujo un principio de prueba basado en la función de score como un método alternativo y con mejores propiedades que el de la razón de la verosimilitud y el método de Wald. (Chandra, 1985). El desarrollo

Methods

Estimation of parameters and overlapping test of Sinha

Sinha *et al.* (2009) developed an overlapping test using the construction method of hypothesis's test developed by Rao (1948). This test used the model (2) and the fact that in this model there is also $\sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, therefore, the logarithm function of reliability comes by

$$l(y; \alpha, \gamma, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \|\mathbf{y} - \alpha \mathbf{X} \gamma\|^2, \quad (3)$$

Where $Q = I + \alpha W$.

On $H_0: \alpha = 0$, the logarithm function of reliability for the random variable has the following estimators

$$\tilde{\gamma} = (\mathbf{X}_*^t \mathbf{X}_*)^{-1} \mathbf{X}_*^t \mathbf{y} \quad (4)$$

y

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}_* \tilde{\gamma})^t (\mathbf{y} - \mathbf{X}_* \tilde{\gamma}), \quad (5)$$

Where \mathbf{X}_* is the designed matrix of the reparameterized model so can be guaranteed the existence of $(\mathbf{X}_*^t \mathbf{X}_*)^{-1}$.

Rao (1948) introduced a test's principle based on the score's function as an alternative method and with better properties than reliability and the Wald's method (Chandra, 1985). The development of the method for the model (2) starts with the definition of (3), in order to obtain the vector's factor.

$$s(\mathbf{y}; \theta) = \partial l(\mathbf{y}; \theta) / \partial \theta \quad (6)$$

With $\theta = (\alpha, \gamma, \sigma^2)$, which is known as score's function. For a θ given and a randomized variable Y the expected value and the variance of (6) are respectively zero and the Fisher's

del método para el modelo (2) comienza con la definición de (3) para posteriormente obtener la función del vector

$$s(\mathbf{y}; \boldsymbol{\theta}) = \partial l(\mathbf{y}; \boldsymbol{\theta}) / \partial \boldsymbol{\theta} \quad (6)$$

con $\boldsymbol{\theta} = (\alpha, \gamma, \sigma^2)$, la cual es conocida como la *función de score*. Para un $\boldsymbol{\theta}$ dado y una variable aleatoria Y tenemos que el valor esperado y la varianza de (6) son respectivamente cero y la matriz de información de Fisher, esta última denotada con $\mathfrak{S}(\boldsymbol{\theta})$.

Para probar la hipótesis nula $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$, el estadístico de Rao viene dado por

$$R_{st} = s(\gamma; \boldsymbol{\theta}_r) \mathfrak{S}^{-1}(\boldsymbol{\theta}_r) s(\gamma; \boldsymbol{\theta}_r), \quad (7)$$

donde $\boldsymbol{\theta}_r$ es el vector de parámetros restringido por la hipótesis nula. Bajo H_0 este estadístico tiene distribución asintótica Chi-cuadrado central con g grados de libertad bajo condiciones generales, donde g representa la dimensionalidad del parámetro de interés (Maddala *et al.*, 1993).

Si la hipótesis que deseamos contrastar es

$$H_0: \alpha = 0 \text{ contra } H_a: \alpha \neq 0, \quad (8)$$

entonces el estadístico score para el contraste en (8) viene dado por

$$RST_0 = s_{\alpha}(0, \hat{\gamma}_r, \hat{\sigma}_r^2) \mathfrak{S}_{\alpha}^{-1}(0, \hat{\gamma}_r, \hat{\sigma}_r^2) s(0, \hat{\gamma}_r, \hat{\sigma}_r^2) \quad (9)$$

donde $\hat{\gamma}_r$ es el estimador restringido para (4), $\hat{\sigma}_r^2$ es la varianza restringida (5) y $\mathfrak{S}_{\alpha}^{-1}$ es el componente de la inversa de la matriz de información esperada de Fisher asociado al parámetro de interés α y evaluada en los estimadores $\hat{\gamma}_r$ y $\hat{\sigma}_r^2$, la cual es no

information matrix is marked as $\mathfrak{S}(\boldsymbol{\theta})$.

To prove the null hypothesis $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the statistics of Rao is

$$R_{st} = s(\gamma; \boldsymbol{\theta}_r) \mathfrak{S}^{-1}(\boldsymbol{\theta}_r) s(\gamma; \boldsymbol{\theta}_r), \quad (7)$$

Where $\boldsymbol{\theta}_r$ is the vector of parameters restricted by the null hypothesis. Under H_0 this statistic has an asymptotic distribution central-squared Chi with g freedom grades under general conditions, where g represents the dimension of the interest's parameter (Maddala *et al.*, 1993).

If the hypothesis that wants to be contrasted is

$$H_0: \alpha = 0 \text{ against } H_a: \alpha \neq 0, \quad (8)$$

Therefore, the statistical scores for the contrast in (8) is represented by

$$RST_0 = s_{\alpha}(0, \hat{\gamma}_r, \hat{\sigma}_r^2) \mathfrak{S}_{\alpha}^{-1}(0, \hat{\gamma}_r, \hat{\sigma}_r^2) s(0, \hat{\gamma}_r, \hat{\sigma}_r^2) \quad (9)$$

where $\hat{\gamma}_r$ is the restricted estimator for (4), $\hat{\sigma}_r^2$ is the restricted variance (5) y $\mathfrak{S}_{\alpha}^{-1}$ is the inverse component of the information matrix expected from Fisher associated to the parameter of interest α and evaluated on the estimators $\hat{\gamma}_r$ y $\hat{\sigma}_r^2$, which is non singular in the reparameterized model. The score function associated to the parameter of interest and already evaluated $(0, \hat{\gamma}_r, \hat{\sigma}_r^2)$ is

$$S_{\alpha}(0, \hat{\gamma}_r, \hat{\sigma}_r^2) = \frac{1}{\sigma_r^2} (Y - X_* \hat{\gamma}_r)' (WX_* \hat{\gamma}_r). \quad (10)$$

The inverse of the information matrix expected $\mathfrak{S}^{-1}(0, \hat{\gamma}_r, \hat{\sigma}_r^2)$ is

$$\hat{\sigma}_r^2 \begin{pmatrix} (\|\hat{H}\|^2 + \|\hat{H}_1\|^2 \hat{H}_1' X_* \hat{C} X_*' H_1) \|\hat{H}\|^2 \|\hat{H}\|^2 \hat{H}_1' X_* \hat{C} & 0 \\ \|\hat{H}_1\|^2 G X_*' H_1 & G & 0 \\ 0 & 0 & \frac{2\hat{\sigma}_r^2}{\gamma} \end{pmatrix} \quad (11)$$

singular en el modelo reparametrizado. La función de score asociada al parámetro de interés y ya evaluada en $(0, \hat{\gamma}_r, \hat{\sigma}_r^2)$ viene dada por

$$s_\alpha(0, \hat{\gamma}_r, \hat{\sigma}_r^2) = \frac{1}{\hat{\sigma}_r^2} (\gamma - X_* \hat{\gamma}_r)' (W X_* \hat{\gamma}_r) \quad (10)$$

La inversa de la matriz de información esperada $\hat{\mathfrak{S}}^{-1}(0, \hat{\gamma}_r, \hat{\sigma}_r^2)$ viene dada por

$$\hat{\sigma}_r^2 \begin{pmatrix} (\|\hat{H}\|^2 + \|\hat{H}_r\|^2 \hat{H}_r' X_* \hat{G} X_*' \hat{H}_r) \|\hat{H}\|^2 \hat{H}_r' X_* \hat{G} & 0 \\ -\|\hat{H}_r\|^2 \hat{G} X_*' \hat{H}_r & G \\ 0 & 0 & \frac{\hat{\sigma}_r^2}{n} \end{pmatrix} \quad (11)$$

con $\hat{H}_r = W X_* \hat{\gamma}_r$, $M_{H_r} = \hat{H}_r \|\hat{H}_r\|^2 \hat{H}_r'$ y $\hat{G} = (X_* (I - M_{H_r}) X_*)^{-1}$. Seleccionando de (11) solamente $\hat{\mathfrak{S}}_\alpha^{-1}$ (expresión dentro de (11)) podemos construir el Test Score de Rao para solapamiento, resultando ser

$$RST_0 = \frac{n \cdot y (I - \tilde{M}) W \tilde{M} y^2}{W \tilde{M} y^2 (I - \tilde{M}) y^2} \cdot \left(1 + \frac{(W \tilde{M} y)' X (X' (I - M_2) X)^{-1} X' (W \tilde{M} y)}{W \tilde{M} y^2} \right) \quad (12)$$

donde $\tilde{M} = X_* (X_* X_*)^{-1} X_*'$ y $M_2 = W \tilde{M} \gamma ((W \tilde{M} \gamma) W \tilde{M} \gamma)^{-1} (W \tilde{M} \gamma)$.

El estadístico (12) posee distribución χ^2 con 1 grado de libertad, correspondiente a la dimensión del coeficiente de solapamiento en el modelo de Draper y Guttman.

Desarrollo de la matriz de pesos en el patrón de siembra del experimento

El experimento consistió en la evaluación del rendimiento (peso en gramos de cada planta) del cultivo de lechuga (*Lactuca sativa* L) sometido a tres niveles de tratamientos y a dos

with $\hat{H}_r = W X_* \hat{\gamma}_r$, $M_{H_r} = \hat{H}_r \|\hat{H}_r\|^2 \hat{H}_r'$ and $\hat{G} = (X_* (I - M_{H_r}) X_*)^{-1}$. Selecting from (11) only $\hat{\mathfrak{S}}^{-1}$ (expression on ()) can be built the Rao test score for overlapping, resulting to be

$$RST_0 = \frac{n \cdot y (I - \tilde{M}) W \tilde{M} y^2}{W \tilde{M} y^2 (I - \tilde{M}) y^2} \cdot \left(1 + \frac{(W \tilde{M} y)' X (X' (I - M_2) X)^{-1} X' (W \tilde{M} y)}{W \tilde{M} y^2} \right) \quad (12)$$

Where $\tilde{M} = X_* (X_* X_*)^{-1} X_*'$ and $M_2 = W \tilde{M} \gamma ((W \tilde{M} \gamma) W \tilde{M} \gamma)^{-1} (W \tilde{M} \gamma)$.

The statistical (12) has distribution χ^2 with liberty degree corresponding to the overlapping coefficient dimension in the Draper and Guttman model.

The development of the weight matrix in the sow pattern of the experiment

The experiment consisted on the yield's evaluation (weight in grams of each plant) of the lettuce's crop (*Lactuca sativa* L) submitted to three treatment levels and two blocks' levels. On this essay was used a line-sow pattern of lettuce with same distance between plants and lines by the convenient at the moment of constructing the weight matrix W . The treatments were labeled as (t_1) the absence of the fertilizer, (t_2) the one which used chemical fertilizer (3 g/plant of the formula 15:15:15) and (t_3) the one which used organic fertilizer (200 g of Gallinanza per plant. The block reason was attributed to the precedence of the commercial seed, where (b_1) represented the seed with national precedence and (b_2) the imported seed of the same species, using this factor to minimize the variability that might be attributable to the precedence of the seed. In the essay were used 5 replications for the combinations of

niveles de bloques. En este ensayo se utilizó un patrón de siembra por hileras de lechuga con igual distancia entre plantas e hileras por lo conveniente al momento de construir la matriz de pesos W . Los tratamientos se etiquetaron como (t_1) la ausencia de fertilizante, (t_2) al que usó fertilizante químico (3 g/planta de fórmula 15:15:15) y (t_3) al que usó el fertilizante orgánico (200 g de Gallinaza por planta). La razón de bloqueo se atribuyó a la procedencia de la semilla comercial, en donde (b_1) representó la semilla de procedencia nacional y (b_2) la semilla importada de la misma especie, usando este factor para minimizar la variabilidad que pudiera ser atribuible a la procedencia de la semilla. En el ensayo se utilizaron 5 repeticiones para las combinaciones de tratamiento y bloque. Cada planta fue considerada como una unidad experimental y se midió como respuesta el rendimiento.

En el modelo (2) y utilizando este ensayo ilustrativo para la aplicación del test, Y es un vector aleatorio que denota la respuesta de $n=30$ unidades experimentales, X_* es una matriz conocida (la reparametrización de la matriz de diseño X), γ es un vector de parámetros desconocidos de dimensión 4×1 (β reparametrizado), w_{ijr} es una matriz de pesos conocida, donde w_{ijr} denota el efecto de la unidad j sobre la unidad i en la réplica r ; $\sum_{j=1}^n w_{ijr} = 1 \forall i, r$; $w_{ijr} = 0$ and $w_{ijr} \geq 0 \forall i, j, r$.

Para el patrón de siembra en hileras de la figura 2, la matriz W se construye una vez se hayan asignado los pesos, por lo cual es necesario conocer la estructura de vecindad que tendrán las diferentes unidades experimentales. En la siembra por hileras

treatments and block. Each plant was considered as an experimental unit and yield was measured as response.

The model (2) and using this illustrative essay for the explanation of the test, is a randomized vector that implies the response of $n=30$ experimental units, X_* is a known matrix (the reparameterized of the design matrix X), γ is a known parameter vector of dimension 4×1 (β reparameterized), $W=(w_{ijr})$ is a known weight matrix, implies the unit effect j on the unit i in the replication r ; $\sum_{j=1}^n w_{ijr} = 1 \forall i, r$; $w_{ijr} = 0$ and $w_{ijr} \geq 0 \forall i, j, r$.

For the line-sow pattern of the figure 2, the matrix W builds once assigned the weights w_{ijr} , therefore, is necessary to know the neighbor's structure that the different experimental units will have. On the line-sow there is an equal price but between plants and lines (w_h, w_v) and a different one between diagonal plants (w_d), all proportional to the sow's distance which varies from producer to producer, also, assuming that only the nearest neighbor ($(h$; among plants), vertical (v ; among lines) and diagonal (d ; among lines)) might be the experimental unit directly affected, will be obtained the neighbor's models shown in figure 1.

The previous pattern differs from the presented by Sinha *et al.* (2009) where are only considered horizontal, vertical and non diagonal neighbors. In this case the consideration is due to the interest on having a non singular weight matrix.

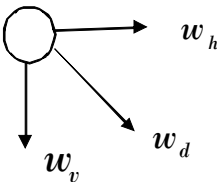
In figure 2 are shown the two sow-lines of lettuce along to the data matrix with the aim of facilitating the weight matrix generated by the

se tiene igual peso entre plantas e hileras (w_h , w_v) y uno diferente entre plantas diagonales (w_d), todos proporcionales a la distancia de siembra que varia de productor a productor, además, asumiendo que solo el vecino más cercano (horizontal (h ; entre plantas), vertical(v ; entre hileras) y diagonal(d ; entre hileras)) puede ser la unidad experimental directamente afectada, tendremos los modelos de vecindad que se muestran en la figura 1.

El patrón anterior difiere del presentado por Sinha *et al.* (2009) en el que solo se consideran vecinos horizontales y verticales y no diagonales. En este caso la consideración se debe al interés en tener una matriz de pesos no singular.

En la figura 2 se muestran las dos hileras de la siembra de lechuga junto con la matriz de datos con la intención de facilitar la construcción de la matriz de pesos generados por la aleatorización de los tratamientos. En cada corchete se presenta el par $\begin{bmatrix} t_t \\ b_b \end{bmatrix}$ que señala al tratamiento t y al bloque b ($t=1,2,3;b=1,2$) para cada réplica en las dos hileras de siembra.

A continuación se ilustrará la construcción de la primera fila de W .



randomized of the treatments. On each bracket are presented the pair $\begin{bmatrix} t_t \\ b_b \end{bmatrix}$ that mention the treatment t and the block b ($t=1,2,3;b=1,2$) for each replication in the two sow lines.

Following will be illustrated the construction of the first line of W . On the response vector γ_{ijr} , the first line corresponds to γ_{111} , which on figure 2 corresponds to the second line and consists on the first treatment, first block and first replication. In this case $\gamma_{111}=135$. This unit has as nearest neighbors the previously mentioned units with the signs. The weights corresponding to the diagrams are obtained from:

$$(a) w_h + w_v + w_d = 1 = w + w + w / \sqrt{2} = 1 = w = (4 - \sqrt{2}) / 7$$

$$(b) 2w_h + w_v + 2w_d = 1 = 2w + w + 2w / \sqrt{2} = 1 = w = (3 - \sqrt{2}) / 7$$

With these values can already be shown the general way of the line number one of the W matrix, since the experimental unit surrounded with signs in the figure 2 corresponds to the model (b) of the neighborhood pattern in figure 1, thus resulting

It can be verified that the addition of the elements of this line is

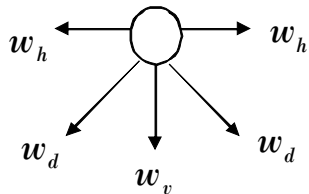


Figura 1. Modelos de vecindad para el cálculo de los pesos en el patrón de siembra en hileras.

Figure 1. Neighborhood model for the calculation of weight in the sow-line pattern.

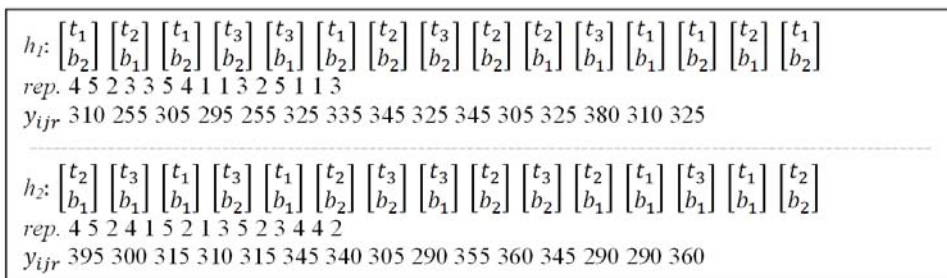


Figura 2. Rendimiento (g.planta⁻¹) en dos hileras (h1,h2) y cinco réplicas (rep.) de los tratamientos.

Figure 2. Yield (g/plant) in two lines (h₁,h₂) and five replications (rep.) of treatments.

En el vector de respuestas y_{ijr} , la primera fila se corresponde con y_{111} , la cual en la figura 2 se corresponde con la segunda hilera y consiste del primer tratamiento, primer bloque y la primera réplica. En este caso $y_{111}=135$. Esta unidad tiene como vecinos más cercanos a las unidades señaladas con las flechas. Los pesos correspondientes a los diagramas se obtienen de:

(a) $w_h + w_v + w_d = 1 = w + w + w / \sqrt{2} = 1 = w = (4 \cdot \sqrt{2}) / 7$

(b) $2w_h + w_v + 2w_d = 1 = 2w + w + 2w / \sqrt{2} = 1 = w = (3 \cdot \sqrt{2}) / 7$

Con estos valores ya podemos mostrar la forma general de la fila 1 de la matriz W , ya que la unidad experimental rodeada de flechas en la figura 2 se corresponde con el modelo (b) del patrón de vecindad en la figura 1, resultando entonces.

$$w_{1jr} = \left(0_{1 \times 9} \frac{(3\sqrt{2}-2)}{14} \ 0_{1 \times 9} \frac{(3-\sqrt{2})}{7} \ 0_{1 \times 2} \frac{(3-\sqrt{2})}{7} \ 0_{1 \times 4} \frac{(3\sqrt{2}-2)}{14} \ \frac{(3-\sqrt{2})}{7} \ 0 \right)$$

Puede verificarse que la suma de los elementos de esta fila es la unidad y que solo existen valores no nulos en los vecinos señalados en la figura 2.

the unit and there are only non null values in the neighbors pointed in figure 2.

Development of the restrictions' matrix

The restriction $\Lambda' \beta = 0$ can be formulated in a finite number of ways,

$$w_{1jr} = \left(0_{1 \times 9} \frac{(3\sqrt{2}-2)}{14} \ 0_{1 \times 9} \frac{(3-\sqrt{2})}{7} \ 0_{1 \times 2} \frac{(3-\sqrt{2})}{7} \ 0_{1 \times 4} \frac{(3\sqrt{2}-2)}{14} \ \frac{(3-\sqrt{2})}{7} \ 0 \right)$$

since $\beta \perp c(\Lambda)$ only if, therefore, if there is any matrix Γ with $c(\Gamma) = c(\Lambda)$, the restriction might be written additionally as $\beta \perp c(\Gamma) \circ \Gamma' \beta = 0$. To identify the reparameterized model was selected the matrix Z thus $c(Z) = c(\Gamma)$, implying that $\Lambda' \beta = 0$ if $\beta \perp c(\Lambda)$ if $\beta \in c(Z)$, which happens if and only if for any vector γ , $\beta = Z\gamma$.

$\Lambda = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ is a two-way classification model with three treatments and two blocks. A non estimable restriction for this model is the one that yields the same lateral conditions in usual experimental units,

that is, $\sum_{i=1}^t \tau_i = 0$ and $\sum_{j=1}^b \delta_j = 0$. A selection of Z that satisfies $c(Z) = c(\Lambda)^\perp$ might be

Desarrollo de la matriz de Restricciones

La restricción $\Lambda'\beta = 0$ puede formularse en un número infinito de formas, ya que $\Lambda'\beta = 0$ si y solo si $\beta \perp \mathcal{C}(\Lambda)$, de modo que si tenemos cualquier otra matriz Γ con $\mathcal{C}(\Gamma) = \mathcal{C}(\Lambda)$, la restricción podría escribirse adicionalmente como $\beta \perp \mathcal{C}(\Gamma)$ o $\Gamma'\beta = 0$. Para identificar el modelo reparametrizado seleccionamos la matriz Z de modo tal que $\mathcal{C}(Z) = \mathcal{C}(\Gamma)$, implicando que $\Lambda'\beta = 0$ si y solo si $\beta \perp \mathcal{C}(\Lambda)$ si y solo si $\beta \in \mathcal{C}(Z)$, lo cual ocurre si y solo si para algún vector γ , $\beta = Z\gamma$.

Sea $\Lambda = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ en un modelo de clasificación de dos vías con tres tratamientos y dos bloques. Una restricción no estimable para este modelo es aquella que rinde las mismas condiciones laterales usuales en este diseño experimental, es decir, $\sum_{i=1}^c \tau_i = 0$ y $\sum_{j=1}^b \delta_j = 0$. Una elección de Z que satisface $\mathcal{C}(Z) = \mathcal{C}(\Lambda)^\perp$ podría ser

$$Z = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & -2 \\ 0 & 0 & -2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

ya que el rango de Z es igual al rango de X y $\Lambda Z = 0$.

Con la matriz Z obtenemos $X_* = XZ$, con lo cual garantizamos la existencia de $(X_*'X_*)^{-1}$ y por ende la no singularidad de la matriz de información esperada.

Retornando al vector de parámetros original tenemos que

Since the Z rank is equal to the rank for X and $\Lambda Z = 0$.

With the matrix Z is obtained $X_* = XZ$, to which is guaranteed the existence $(X_*'X_*)^{-1}$ thus the non singularity of the information matrix expected.

Returning to the original

$$Z = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & -2 \\ 0 & 0 & -2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & 1 \end{pmatrix},$$

parameters' vector

Which takes to $\tau_1 + \tau_2 + \tau_3 = 0$ and $\delta_1 + \delta_2 = 0$.

Results

The maximum-accuracy estimation of the reparameterized model with the contrast of hypotheses

$$\beta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \delta_1 \\ \delta_2 \end{pmatrix} = Z\gamma = Z \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} = \begin{pmatrix} \gamma_0 - \gamma_2 \\ -\gamma_1 + \gamma_2 - 2\gamma_3 \\ -2\gamma_2 + \gamma_3 \\ \gamma_1 + \gamma_2 + \gamma_3 \\ \gamma_1 - \gamma_3 \\ -\gamma_1 + \gamma_3 \end{pmatrix}$$

given in (8), that is, null hypotheses of the “overlapping absence” yields the estimated parameters and, which are obtained thanks to the use of a simple algorithm done with the statistical package SAS, on the module SAS/IML and the appropriate substitution of the data presented in figure 2.

Of the vector $\hat{\gamma}$ is also obtained that

The Rao score test for overlapping evaluated on the estimators maximum-accuracy of the

$$\beta = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \delta_1 \\ \delta_2 \end{pmatrix} = Z\gamma = Z \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} = \begin{pmatrix} \gamma_0 - \gamma_2 \\ -\gamma_1 + \gamma_2 - 2\gamma_3 \\ -2\gamma_2 + \gamma_3 \\ \gamma_1 + \gamma_2 + \gamma_3 \\ \gamma_1 - \gamma_3 \\ -\gamma_1 + \gamma_3 \end{pmatrix},$$

lo cual conduce a $\tau_1 + \tau_2 + \tau_3 = 0$ y $\delta_1 + \delta_2 = 0$.

Resultados

La estimación máximo-verosímil del modelo reparametrizado con el contraste de hipótesis dado en (8), es decir, la hipótesis nula de “ausencia de solapamiento” rinde los parámetros estimados $(321.83 \quad -1.40 \quad 0.47 \quad 7.10)$ y $\hat{\sigma}^2 = 850.83$

los cuales se obtienen gracias al uso de un simple algoritmo hecho con el sistema estadístico SAS, dentro del módulo SAS/IML y a la apropiada sustitución de los datos presentados en la figura 2.

Del vector $\hat{\gamma}$ también se obtiene que

$$\begin{aligned} \hat{\beta}_0 &= 321.83 = \widehat{\gamma_0 - \gamma_2}, \\ \hat{\beta}_1 &= -12.33 = -\widehat{\gamma_1 + \gamma_2 - 2\gamma_3}, \\ \hat{\beta}_2 &= 6.17 = -\widehat{2\gamma_2 + \gamma_3}, \\ \hat{\beta}_3 &= 6.17 = \widehat{\gamma_1 + \gamma_2 + \gamma_3}, \\ \hat{\beta}_4 &= -8.5 = \widehat{\gamma_1 - \gamma_3}, \\ \hat{\beta}_5 &= 8.5 = -\widehat{\gamma_1 + \gamma_3}. \end{aligned}$$

El Score Test de Rao para solapamiento evaluado en los estimadores máximo-verosímiles del modelo reparametrizado resultó en 5.081, el cual comparado con la distribución Chi-cuadrado con un grado de libertad genera una significación de 0.0242 (2.42%), con lo cual se rechaza la hipótesis nula de ausencia de

reparameterized model resulted in 5.081, which compared to the squared-Chi distribution with a freedom degree generates a significance of 0.0242 (2.42%), to which is rejected a null

$$\begin{aligned} \hat{\beta}_0 &= 321.83 = \widehat{\gamma_0 - \gamma_2}, \\ \hat{\beta}_1 &= -12.33 = -\widehat{\gamma_1 + \gamma_2 - 2\gamma_3}, \\ \hat{\beta}_2 &= 6.17 = -\widehat{2\gamma_2 + \gamma_3}, \\ \hat{\beta}_3 &= 6.17 = \widehat{\gamma_1 + \gamma_2 + \gamma_3}, \\ \hat{\beta}_4 &= -8.5 = \widehat{\gamma_1 - \gamma_3}, \\ \hat{\beta}_5 &= 8.5 = -\widehat{\gamma_1 + \gamma_3}. \end{aligned}$$

hypothesis of overlapping absence for a trustable level of 95%.

It is important to highlight that any other matrix that fulfill the specifications mentioned in the latter section, will yield a low equivalent model (8) but with different parametrization and in both cases will fulfill the lateral imposed conditions.

Conclusions

It has been illustrated carefully the statistical overlapping test based on the Rao test score using operations with perpendicular projections and non estimable functions developed by Sinha *et al.* The test requires that the structure of the experimental design involves a model like the one of Draper and Guttman, using as entrance variables the different factors of two-way classification model.

The asymptotic nature of the test requires the use of big-sized samples, for such reason it is very useful in the agronomic and animal production areas, since these areas normally work with a big replication number of the experimental units, which might

solapamiento para un nivel de confianza del 95%..

Es importante destacar que cualquier otra matriz que cumpla las especificaciones señaladas en la sección inmediatamente anterior, rendirá un modelo equivalente bajo la hipótesis (8) pero con diferente parametrización y en ambos casos se cumplirán las condiciones laterales impuestas.

Conclusiones

Se ha ilustrado detalladamente la prueba estadística para el solapamiento basado en el test Score de Rao con el uso de operadores de proyección perpendicular y funciones no estimables desarrollada por Sinha *et al.* La prueba requiere que la estructura del diseño experimental involucre un modelo como el de Draper y Guttman utilizando como variables de entrada los diferentes factores de un modelo de clasificación en dos vías.

La naturaleza asintótica del test requiere del uso de tamaños de muestras grandes, por tal motivo suele ser bastante útil en el área de la agronomía y en la producción animal, pues en dichas áreas suele trabajarse con un número grande de repeticiones de las unidades experimentales, lo cual probablemente asegurará el cumplimiento de los supuestos necesario para el análisis.

Es importante reconocer la dificultad asociada en la asignación de pesos de la matriz, solo en el ejemplo de aplicación fue necesaria una matriz de dimensión, la cual no puede sistematizarse pues la asignación de los tratamientos es totalmente aleatoria y su con-

assure the fulfillment of the suppose necessary for the analysis.

It is important to know the difficulty associated in the assignation of weights of the W matrix, only on the application example was necessary a dimension matrix 30×30 , which cannot be systematized because the assignation of treatments is totally at random and the conformation depends on each randomized activity.

Finally, and though the essay done in the field was just used to illustrate the application of the test, the detection of overlapping presence may suggests the presence of external factors that might significantly scatter the treatments or part of these in the closest units. The overlapping absence, from the model point of view, establishes as definite model the non lineal model (2), thus, the comparison technique of treatments posterior to the variance analysis might be based on the non lineal model (1), on the non lineal reparameterized model (2) or just on an usual lineal model where is known the overlapping coefficient value.

End of english version

formación depende de cada aleatorización.

Finalmente, y aunque el ensayo realizado en campo fue solo para ilustrar la aplicación del test, la detección de presencia de solapamiento sugiere probablemente la presencia de factores externos que pudieran significativamente dispersar los tratamientos o parte de estos en las unidades más próximas. La ausencia de

solapamiento desde el punto de vista del modelo establece como modelo definitivo al modelo no lineal (2), de modo que la técnica de comparación de tratamientos posterior al análisis de varianza pudiera basarse en el modelo no lineal (1), en el modelo no lineal reparametrizado (2) o simplemente en un modelo lineal usual donde se conozca el valor del coeficiente de solapamiento.

Literatura citada

- Bhali, M. A., Day, A. D., Tucker, H., Thomson, R. K., and Massey, G. D. 1964. End-border effects in irrigated barley yield trials. *Agronomy journal* 1956, 346-348.
- Chandra, T. K. and Mukerjee, R. 1985. Comparison of the likelihood ratio, Wald's and Rao's tests. *Sankhyā Ser. A* 47, 271-284.
- Christensen, R. 2002. *Plane Answers to Complex Questions. The Theory of Linear Models*. Third edition. Springer-Verlag New York, Inc.
- Darghan, A. E., Sinha, S. P. y Goitia, A. A. 2011. Score Test para el efecto del coeficiente de solapamiento en modelos de superficie de respuesta de primer y segundo orden. *Revista de la Facultad de Minas de la Universidad Nacional de Colombia, Dyna*. Número 165, año 78, 234-245.
- Draper, N. R. and Guttman, J. 1980. Incorporating overlap effects from neighboring units into response surface models. *Applied Statistics* 39, 128-134.
- Hide, G. A. and Read, P. J. 1990. Effect of neighboring plants on the yield of potatoes from seed tubers affected with gangrene (*Phoma foveata*) or from plants affected with stem canker (*Rhizoctonia solani*). *Annals of Applied Biology* 116, 233-243.
- Maddala, G.S., Rao, C.R., Vinod, H.D. 1993. *Handbook of Statistics*. Vol. 11. North Holland, Amsterdam. 800 pp.
- Pearce, S. C. 1957. Experimenting with organisms as blocks. *Biometrika* 44, pp. 141-149.
- Rao, C. R., 1948. Large sample tests of statistical hypothesis concerning several parameters with applications to problems of estimation. *Proc. Cambridge Philos. Soc.* 44,50-57.
- SAS Institute, Inc. 2004. *SAS/IML 9.1 user's guide*. SAS Inst., Inc., Cary, NC.
- Sinha, S.P., Darghan, A.E. y Goitia, A.A. 2010. A Score Test for overlap effect in the two-way classification model under side conditions on the model parameters. (Unpublished manuscript but accepted as a key outcome of the proposed second's author doctoral thesis in statistics).
- Shukla, G.K., Subrahmanyam, G.S.V. 1999. A Note on a test and Confidence Interval for Competition and Overlap Effects. *Biometrics*, Vol. 55, No. 1, pp. 273-276.