Darghan, Aquiles; Surendra, Sinha; Monroy, Julio
A score test for the agronomical overlap effect in a two-way classification model
Universidad Nacional de Colombia
Bogotá, Colombia

Available in: http://www.redalyc.org/articulo.oa?id=180333786016
A score test for the agronomical overlap effect in a two-way classification model

Aquiles Darghan¹, Sinha Surendra², and Julio Monroy³

ABSTRACT

In some agricultural research, a treatment applied to an experimental unit may affect the response in the neighboring experimental units. This phenomenon is known as overlap. In this article, a test to evaluate this effect in the Draper and Guttman model was developed by imposing side conditions on the parameters of a two-way classification model to obtain a re-parameterized model which can be used in different neighboring patterns of experimental units, usually plants within a crop, whenever the nearest neighbor is considered a directly affected experimental unit and the two-way model is used. Three methods, namely maximum likelihood, least squares with side conditions and generalized inverse, were used to estimate the parameters of the original model in order to calculate the value of the test statistics for the null hypothesis associated with the absence of the overlapping effect. The three alternatives were invariant with respect to the use of test. The proposed test is simple to adopt and can be implemented in agronomy since its asymptotic nature is in agreement with the large number of experimental units which generally exist in this type of research, where each plant represents the experimental unit being assessed.

Key words: perpendicular projection operator, side conditions, experimental design.

Introduction

In some agricultural tests, a treatment that is applied to an experimental unit may affect the response in the neighboring experimental units. This phenomenon is known as overlap. For example, in variety testing, the effect of a neighbor can be attributed to differences in height between plants, strength of roots, and date of germination, among others. Hide and Read (1990) discussed this situation in potato cultivation. The treatments applied to crops for fertilization plans, irrigation and pesticide applications can be dispersed to other plots or adjacent experimental units, which can affect the response in the neighboring units. An example of this situation for an irrigation experiment can be found with Bhalli et al. (1964). This overlapping phenomenon has been modeled by several researches; for example, Pearce (1957) considered a model in which each treatment had a direct influence on the plot in which they were applied and an overlap effect on the neighboring plots. Draper and Guttman (1980) also considered such a
model, discussing some approximate testing methods as well as a confidence interval for the overlap effect. Draper and Guttman (1980) used a nonlinear model in which the overlapping was attributed to all the effects considered in the model. Such a model is written as:

\[ Y = X\beta + \alpha W X \beta + \epsilon \]  

(1)

where the random vector \( Y \) denotes the response of \( n \) experimental units, \( X \) is a known design matrix of dimension \( n \times p \) with rank \( q > p \), \( \beta \) is a \( p \)-dimensional vector of unknown parameters consisting of the effects of blocks and treatments, \( \alpha \) is the overlapping coefficient, and \( W = (w_{ij}) \) is a matrix of known weights of dimension \( n \times n \), where \( w_{ij} \) denotes the effect of unit \( j \) on unit \( i \); such that \( \sum_{j=1}^{n} w_{ij} = 1 \) for all \( i \); and \( w_{ii} = 0, w_{ij} \geq 0 \) for all \( i \) and \( j \). The fact that the matrix is of an incomplete rank generates the possibility of imposing side conditions on the parameters so that an estimator for the \( \beta \) vector can be obtained. Finally, it is assumed that the distribution of the error vector \( \epsilon \) is normal and independent with an expected value of zero (\( E(\epsilon) = 0 \)) and variance of \( \sigma^2I \), where \( I \) is an identity matrix. Shukla and Subrahmanyan (1999) considered a generalization of the first model represented by (1), which only included a subset of all the direct effects influencing the neighboring plots. In the following sections of the article, some important results which allowed the construction of Rao’s score test for \( \alpha \) of model (1) are presented; in the methods section, we explain the side condition technique, the obtainment of several matrices associated with the model of Draper and Guttman (1980) and, finally, we present Rao’s score test under the imposition of side conditions.

Methods

Side conditions and associated matrices in the Draper and Guttman model

The technique to impose side conditions is very well-known in the area of experimental design since it provides the necessary (linear) restrictions which can assure that the estimation of the parameters is unique. Another use of side conditions is the imposition of arbitrary restrictions on the estimators so that the normal equations can be simplified. In this case, the estimators have exactly the same behavior when a generalized inverse of \( X^TX \) is used to obtain the vector of parameters. Let \( q \) be the rank of the design matrix \( X \) where \( q < p < n \) in (1), then \( X^TX \beta \) represents a set of \( p \) estimable functions of \( \beta \). The side conditions must be non-estimable functions of \( \beta \). Since the rank of the \( X \) is \( q \) and, hence, the deficiency of this rank is \( p-q \), conditions such as \( L \beta = 0 \) or \( L \hat{\beta} = 0 \) to obtain a unique solution for \( \hat{\beta} \) should be defined where \( L \) is a matrix of dimension \( (p-q) \times p \) with a rank of \( L \) being \( p-q \) such that \( L \beta \) is a set of non-estimable functions. The (1) model under the null hypothesis \( \alpha = \alpha_0 \) along with the side conditions is written as:

\[
\begin{bmatrix}
Y_{nl} \\
Y_{0l}
\end{bmatrix} = \begin{bmatrix}
X_{nxp} & \alpha_0 \end{bmatrix} \begin{bmatrix}
\beta_{p1} \\
\alpha_0
\end{bmatrix} + \begin{bmatrix}
W_{nxx} & 0_{n0} \\
0_{00} & 0_{00}
\end{bmatrix} \begin{bmatrix}
X_{0xp} \\
0_{0l}
\end{bmatrix} \beta_{p1} + \begin{bmatrix}
\epsilon_{n1} \\
0_{0l}
\end{bmatrix}
\]

or

\[ Y_i = U \beta + \alpha_0 W_i U \beta + \epsilon_i \]

(2)

where \( U^T = [X^T L^T] \) has rank \( p \), \( W_i \) is a matrix of weights with \( l \) additional rows, and where \( l \) represents the number of imposed linear restrictions. The \( U^T U \) matrix is of a \( p \times p \) dimension and rank \( p \), which generate the system of normal equations \( U^T U \hat{\beta} = U^T Y \), where \( Y \) is the vector of responses expanded with zeros and has a unique solution for \( \hat{\beta} \) (Graybill, 1976). In order to illustrate the above described procedure, let us suppose that we have a two-way classification model (treatments and blocks) which will be written as \( Y_{ij} = \mu + \tau_i + \delta_j + e_{ij} \) for \( i = 1, 2, ..., t \), \( j = 1, 2, ..., b \) with \( \tau_i \) as the effect of the \( i \)th treatment and \( \delta_j \) as the effect of the \( j \)th block, which in matrix form is written as \( Y = X \beta + \epsilon \). In this model, there is a deficiency in rank \( p-q = 2 \), with \( p = t + b + 1 \) and \( q = t + b - 1 \). In order to solve this problem of rank deficiency in this design model, the usual non-estimable functions \( \sum_{i=1}^{t} \tau_i = 0 \) and \( \sum_{j=1}^{b} \delta_j = 0 \) can be used, which can be expressed in the matrix form as the following set of non-estimable functions:

\[
L \beta = \begin{bmatrix}
0 \\
J^T_t \\
0 \\
0 \\
0 \\
J^T_b
\end{bmatrix} \beta = \begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

denoting \( J_t \) and \( J_b \) as the vectors of ones of the \( t \) and \( b \) elements, respectively. The matrices \( U^T U, (U^T U)^{-1} \), \( (U^T U)^{-1} X^T \) and \( X(U^T U)^{-1} X^T \) are needed in the estimation of the vector of parameters \( \beta \), error variance and, therefore, for obtaining Rao’s score test for overlapping. However, here, we only present the final expression (Darghan, 2010); it is the comprehensive development of related matrix calculations, namely:

\[
X(U^T U)^{-1} X^T = b^{-1} X_i X_j^T + t^{-1} X_j X_j^T - n^{-1} J J^T = M_1 + M_2 - M_3
\]

(3)

with \( n = bt \), \( X = (J X_i X) \), where \( X_i \) and \( X \) are the submatrices of the design matrix associated with the effects of the treatments and blocks respectively, \( M_1 = b^{-1} X_i X_j^T \), \( M_2 = t^{-1} X_j X_j^T \) and \( M_3 = n^{-1} J J^T \), which are all perpendicular projection matrices onto the column spaces of \( X_i \), \( X \).
and J, respectively, denoted with C(Xi), C(X0) and C(J).
This last statement is shown in a later theorem. The space
spanned by the columns of a vector or matrix called a
column space is written as C( ), placing in the brackets
the vector or matrix of interest. Another interesting use of
side conditions for obtaining the re-parameterization of a
model can be described as follows: Let Lβ = 0 define the
side conditions in the model (1). If Lβ = 0 holds, then β
belongs to the orthogonal (⊥) complement of C(L), namely
β ∈ C(L⊥). Let Z be such that C(L⊥) = C(Z), then β = Zγ
for some γ; then, by substituting β with Zγ and defining
X0 = XZ in model (1), we get

\[ Y = X_{0}\gamma + \alpha WX_{0}\gamma + \varepsilon \]  (4)

Besides, if Lβ is not estimable, then C(X) = C(X0) and the
third model represented by (4) is a re-parameterization of
model (1) (Christensen, 2011).

Rao’s score test
This section summarizes the theory of likelihood for the
score test. Further information about score tests can be
found in (Rao, 1973) and (Cox and Hinkley, 2000). The
article published by (Rao, 1948) introduced the funda-
mental principle of a test based on the score function as
an alternative method to the likelihood ratio test and to
Wald’s method. Several authors have described the at-
tractive properties of the method; among them, Chandra
and Joshi (1985), Bera and Mckenzie (1986) and Bera and
Bilias (2001). For a better understanding of the test, we
introduce its notation, which is maintained until the de-
velopment of the test for overlapping. Suppose we want to test the hypothesis
H0: α=0 against
H1: α≠0 in model (1), under H0, this model is the usual
linear model Y = Xβ + ε; however, X is not of a full column
rank; in this sense, the p parameters in β are not unique.
We then ascertained whether β could be estimated. Using
least-squares in (1), we obtained the normal equations XTXβ
= XTY. However, since X was not of a full column rank, XTX
had no inverse; as a consequence, the normal equations did
not have a unique solution, despite the fact that this system
was consistent if and only if XTX(XTX)−1XTY = XTY, where
(XTX)−1 is a generalized inverse of (XTX). Since the normal
equations were consistent, a solution is:

\[ \hat{\beta} = (X'X)^{-1}X'Y \]  (5)

For a particular generalized inverse (XTX)−1, the expected
value of \( \hat{\beta} \) in (1) under H0 is E(\( \hat{\beta} \)) = (XTX)−1XTYβ, thus \( \hat{\beta} \)
is not an unbiased estimator of β; furthermore, the expression
(XTX)−1XTYβ is not invariant to the choice of (XTX)−1, thus,
\( \hat{\beta} \) in (5) does not estimate β. With respect to the estimation of σ2, in model (1), we define:

\[ \hat{\sigma}^2 = Y^T \left[ I_n - X(X'X)^{-1}X^T \right] Y / (n - q) \]  (6)

where n is the number of rows of X which has rank q. The estimator in (6) is unbiased for σ2 and invariant to the choice of \( \hat{\beta} \) and to the choice of generalized inverse (XTX)−1. For the non-full column rank model (1), under H0, we now
assume that ε is distributed as N(0, σ2I). With the nor-
mality assumption we can obtain a maximum likelihood estimator for (1) under H0. In this case, we have the same
estimator for β which is given by the least square estimation
but a biased estimator for σ2 is obtained, which can be written
as \( \hat{\sigma}^2 = [n \cdot Y^T(1 - X(X'X)^{-1})X^TY] \). The lack of uniqueness in the estimate for \( \beta \) using two different estimation methods

Darghan, Surendra, and Monroy: A score test for the agronomical overlap effect in a two-way classification model
in (1) leads us to the use of (2). In this model, the problem of the rank of $X$ is solved with the imposition of side conditions to proceed with the estimation of the parameters using least squares. Using model (2) restricted by $H_i$ we obtain
\[ \hat{\beta} = (U^T U)^{-1} U^T Y, \]
and
\[ \hat{\sigma}^2 = Y^T [I_n - X(U^T U)^{-1} X^T] Y / (n - q), \]
which are unique and unbiased. Also the error vector has two interesting features: it has singular normal distribution and, therefore, the log-likelihood function doesn’t exist and contains the matrix given in (3), which is very important in the development of the theorems presented below and can be extended in the balanced two-way classification model with more than one observation per cell by just modifying the design and weight matrices in order to adjust for the replications of each treatment.

**Theorem 1.** Let $M_i = b^T X_i X_i^T$ and $M_j = n^{-1} J J^T$ be perpendicular projection matrices onto the respective column spaces of $X_i$, $X_j$ and $J$. To prove that $C(M) = C(X)$, $C(M) = C(X)$, and $C(M) = C(J)$.

**Proof.** An inspection of the matrices $M_i$, $M_j$, and $M_j$ allows for a verification of the symmetry and idempotent properties, which is enough to prove that these matrices are all perpendicular projection matrices onto the respective column spaces of $X_i$, $X_j$, and $J$. To prove that $C(M) = C(X)$, we can refer to the B.51 proposition in (Christensen, 2011). In the cases where $C(M) = C(X)$, $C(M) = C(J)$, the same procedure can be followed.

**Theorem 2.** For matrices $M_i$, $M_j$, and $M_j$, it holds that if $M_i M_j = M_j M_i$, $M_j M_j = M_j$, and $M_i M_j = M_j$, then $C(I - X(U^T U)^{-1} X^T) = C(I - M) \subset C(M)$.

**Proof.** The proof is obvious by the very definition of the matrices $M_i$, $M_j$, and $M_j$ in the two-way classification design without interaction. To prove that $C(I - M) \subset C(M)$, it is sufficient to multiply $I - M$ with $M_j$ and observe that their product is zero. In addition, the rank of $C(M)^{(n-1)}$ as $n - 1$ is greater than the rank of $C(I - M)$, whose value is $n - q$, whenever $q > 1$ in the two-way classification model without interaction.

**Theorem 3.** $M$ is the perpendicular projection matrix on $C(X)$, where $M = X(U^T U)^{-1} X^T$.

**Proof.** In order to prove that $M$ is the perpendicular projection matrix on $C(X)$, it is necessary to verify that $M$ is symmetric and idempotent. The symmetry obviously results from equation (3) when verifying that $M = M^T$. So we only have to prove idempotence ($M = M^2$). Since $M = M_i + M_j - M_j$, therefore $M^2 = (M_i + M_j - M_j)^2$, and using the previous theorems, we have $M = M^2$.

**Theorem 4.** The matrix $(U^T U)^{-1} = (X^T X + L^T L)^{-1}$ is a generalized inverse of $X^T X$.

**Proof.** A generalized inverse of $X^T X$ is any $G$ matrix such that $X^T G X = X^T X$. Let $G = (X^T X + L^T L)^{-1}$, then $X^T G X = X^T X + L^T L$, but replacing $M$ with theorem 3, we have $X^T G X = X^T X$ and as $M X = X$ for the same theorem, we have $X^T G X = X^T X$.

**Theorem 5.** The matrices $M_a = M_i - M_j$ and $M_b = M_i - M_j$ are perpendicular projection operators; additionally, $C(M_a) \perp C(M_b)$, $C(M_b) \perp C(M_a)$ and $C(M_a) \perp C(M_b)$.

**Proof.** To start, we will prove that $M_a$ and $M_b$ are idempotent matrices, which we obtain by verifying that $(M_i - M_j)^2 = M_i^2 + M_j - M_i M_j = M_i - M_j$. The symmetry of $M_i - M_j$, obviously results when we substitute $M_i$ with $b^T X_i X_j^T$ and $M_j$ with $n^{-1} J J^T$, from which we obtain $M_a^2 = M_a$. The same procedure may be used in the case of $M_b$. In order to prove that $C(M_a) \perp C(M_b)$ or, in other words, to prove that $C(M_a)$ is contained in the orthogonal complement of $C(M_b)$, it should be sufficient to verify that the product of the matrices associated with these spaces is null. In the first case, we have $M_i - M_j = M_i - M_j = M_i - M_j + M_j - M_j = 0$; for the cases $C(M_a) \perp C(M_b)$ and $C(M_b) \perp C(M_a)$ the same procedure is followed.

To develop Rao’s score test for overlapping, we need the expected information matrix which is constructed from the log-likelihood function. In the (1) model, the log-likelihood is singular due to the deficiency in the rank of $X$. Thus, models (1) and (2) are not used in the construction of the test, but we still use their estimates for $\beta$ and $\sigma^2$. The log-likelihood function for (4) is:
\[
\ell(\alpha, \gamma, \sigma^2) = -n \log(2\pi) / 2 - n \log(\sigma^2) / 2 - \|Y - X_\alpha \gamma + \sigma W X_\gamma\| / 2
\]
from which, under $H_0$, we obtain:
\[
\hat{\gamma} = (X_\alpha X_\alpha)^{-1} X_\alpha Y
\]
and
\[
\hat{\sigma^2} = n^{-1} Y^T [I_n - M_i] Y
\]
as the perpendicular projection matrix on $C(X_\alpha)$. The score vector is:
\[
s(\alpha, \gamma, \sigma^2) = \frac{1}{\sigma^2} \left( \begin{array}{c} (WX_\gamma)^T (Y - Q_\gamma) \\ X_\alpha^T (Y - Q_\gamma) \\ n / 2 + \|Y - Q_\gamma\| / 2 \end{array} \right)
\]
with $Q = (I_n + \alpha W) X_\alpha$. If we denote $s(\alpha, \gamma, \sigma^2)$ the component of the score vector corresponding to the parameter of interest ($\alpha$), the score statistic to test $H_0: \alpha = 0$ is:

Agron. Colomb. 32(3) 2014
\[ S_{\text{ovp}} = s_n(0, \hat{\gamma}, \hat{\sigma}^2)^\top \hat{\delta}^{-1}(0, \hat{\gamma}, \hat{\sigma}^2) s_n(0, \hat{\gamma}, \hat{\sigma}^2) \]

where \( \hat{\gamma} \) and \( \hat{\sigma}^2 \) are the maximum likelihood estimators of \( \gamma \) and \( \sigma^2 \) in (4) and \( \hat{\delta}(0, \hat{\gamma}, \hat{\sigma}^2) \) is the component in the expected information inverse matrix evaluated at \( \alpha = 0 \). The expected information inverse matrix is:

\[
\hat{\delta}^{-1}(0, \hat{\gamma}, \hat{\sigma}^2) = \begin{bmatrix} H & X & G \vert H \end{bmatrix}^\top \hat{H}^{-1} \begin{bmatrix} H & X & G \vert H \end{bmatrix}^{\top} - \begin{bmatrix} H \vert X \end{bmatrix} \begin{bmatrix} H \vert X \end{bmatrix}^\top \end{bmatrix} \begin{bmatrix} H \vert G \end{bmatrix} 0 \\
0 & G & 0 \\
0 & 0 & 2\hat{\sigma}^2 / n
\]

where \( \hat{H} = WX_nY \) and \( \hat{G} = (X_n^\top(I_n - M_n)X_n)^{-1} \) with \( M_n = \hat{H}[\hat{H}]^{-1} \hat{H}^\top(M_n \text{ as the perpendicular projection operator on } C(H)) \). Now, when evaluating \( s_n(0, \hat{\gamma}, \hat{\sigma}^2) \) in \( s_n(0, \hat{\gamma}, \hat{\sigma}^2) \) and grouping terms, we obtain:

\[ S_{\text{ovp}} = \hat{\sigma}^2 \begin{bmatrix} H \vert H \end{bmatrix}^{-1} H \begin{bmatrix} I_n + X_n \hat{G}X_n^\top \end{bmatrix} \hat{H} \begin{bmatrix} M_{\beta} \vert I_n - M_n \end{bmatrix} Y^\top \] (7)

The score statistic test for overlap has a \( \chi^2 \) distribution with \( g = 1 \) degree of freedom, which tests \( \alpha \) in the Draper and Guttman model. The statistic in (7) depends on \( \hat{\beta} \) only through \( \hat{H} \) (and hence \( \hat{G} \)). The estimates for \( \beta \) and \( \sigma^2 \) in models (1) and (4) can be substituted in (7), verifying that the same value of the statistic is obtained. Such a result is not surprising given that models (1) and (4) have the same estimation space \( C(X) \). Darghan (2010) showed the invariance property of \( S_{\text{ovp}} \) using \( (X^2X + L^2L)^{-1} \) or \( (X^2X)^{-1} \).

**Discussion**

A statistical test based on Rao’s score test has been developed for the overlapping effect in a re-parameterization of the Draper and Guttman model (Draper and Guttman, 1980). Although the phenomenon of overlap is not new in agricultural sciences, evaluation by means of a model is relatively new. Few authors have modeled this phenomenon. In 1999, Shukla and Subrahmanyan (1999) proposed an exact test and confidence intervals to assess the overlap coefficient using the model of Draper and Guttman, but using the Likelihood ratio test. Despite the side conditions imposed in the original model to resolve the no-full-rank restriction in the design matrix and obtain unique and unbiased estimated parameters in (2), the error vector introduced the singular normal distribution; so, in this case, it is not defined as a log-likelihood function and, hence, it was impossible to obtain the estimators by maximum likelihood as it was not possible to apply the methodology of Rao to build the hypothesis testing. So far, we can say that the original model and the extended model by side conditions presented the respective restrictions of not being able to obtain the inverse of the expected Fisher information matrix or singular normal distribution. Thus, the test was developed by re-parameterization of the original model. The natural maximum likelihood estimators were obtained for the parameters \( \gamma \) and \( \sigma^2 \) under \( H_0 \), and, with these same estimators, the score test for overlapping was evaluated. The dependence of the statistical test on the parameters allows for replacing the two sets of estimators obtained by least square estimation: one for \( \beta \) and another for \( \sigma^2 \) in each studied model. The statistical test was invariant, which was expected because models (1) and (4) were equivalent and model (2) provided an estimate for \( \beta \), belonging to the same estimation space of models 1 and 4. The statistical test for overlapping should be of wide application in agronomical research and has extensive practical value because the application of a variance analysis on a data set that was collected in the field has experienced the effect of overlap, generating spurious results as the effect of the treatments is confused since the same experimental unit could be receiving more than one treatment at a time. This test is easy to adopt as long as the layout of the used experimental design involves a model similar to that of Draper and Guttman whenever the nearest neighbor is considered a possible source of overlapping. Although the used example involved a test of overlapping by means of only the effects included in the two-way model, it can also be tested for any number of factors and their interactions as well as for a subgroup of effects and not only the use of the nearest neighbor by just modifying the design matrix and the matrix of the weights in the re-parameterized model. Statistics have been tested in various agricultural applications (Darghan et al., 2012) as well as in the area of education (Darghan et al., 2014) not only in design models, but also in the context of response surface modeling (Darghan et al., 2011), obtaining similar results as observed in the field (Darghan, 2010).

**Conclusions and recommendations**

The developed overlap test can be used in the field of agronomy where it is increasingly suspected that applied treatments can move from one experimental unit to that of the nearest neighbor and that the presence of overlap may invalidate the comparison of treatments when using the analysis of variance associated with a linear model, in this case, a two-way classification model. The asymptotic nature of the test requires a large amount of experimental units for it to be valid. The results obtained in applications using linear models have been similar to those observed in field results. Once the overlap coefficient has been estimated, the analysis of variance can be corrected by the overlap effect. Monte Carlo simulation studies as well as agricultural applications in the field of information technology...
and communication will complement the properties of the test, which can be extended to more complex experimental designs using neighboring patterns that cover all of the experimental units that are being studied.

**Literature cited**


