

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1991 - 3rd Annual Conference Proceedings

EVALUATION OF FOUR COVARIATE TYPES USED FOR ADJUSTMENT OF SPATIAL VARIABILITY

Paul N. Hinz

John P. Lagus

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Hinz, Paul N. and Lagus, John P. (1991). "EVALUATION OF FOUR COVARIATE TYPES USED FOR ADJUSTMENT OF SPATIAL VARIABILITY," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1422>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

EVALUATION OF FOUR COVARIATE TYPES USED FOR ADJUSTMENT OF SPATIAL VARIABILITY¹

Paul N. Hinz and John P. Lagus
Department of Statistics
Iowa State University

Key Words: Field-plot experiments, Nearest-neighbor analysis, spatial statistics.

Abstract

Four types of covariates are used to account for spatial variability in data from a field experiment for evaluating 620 soybean varieties for iron chlorosis. The covariates are calculated as the average of 4 and of 14 neighboring residuals and of 4 and of 14 neighboring observations. The residual mean square from the analysis of covariance was smaller when residuals were used in calculation of the covariates than when observations were used. Moreover, use of 14 neighbors resulted in smaller residual mean squares than did use of 4 neighbors. Differences among 4 covariate types were small and not practically important. Expected values for the covariate regression coefficients were derived based on an errors in variables model. The expected values depend only on the measurement error of the covariate and are unrelated to the strength of the spatial variability. The coefficients estimated from the analysis of covariance are generally greater than the expected values.

1. Introduction

Randomized block designs are used in field experiments to reduce the experimental error due to spatial variability. When spatial variability cannot be satisfactorily controlled by blocks, it would be desirable to have a method of analysis that would account for this variability. The purpose of this paper is to demonstrate the use of a nearest-neighbor type of analysis and the possible limitations of interpretations of the analysis.

The data which are used to illustrate the analysis come from a randomized block design with 4 blocks and 620 treatments. The treatments were soybean varieties planted in single row plots 3' long and 18" apart. Each block had plots arranged in an array of 31 rows and 20 columns.

The experiment was conducted to evaluate the varieties for resistance to iron chlorosis. Iron chlorosis is a condition in which plants cannot extract iron from the soil. This condition occurs on soil of high pH and is detected by yellowing of the leaves. The amount of chlorosis in each plot was measured by scoring the amount of yellowing on a scale of 1 to 5 in .5 steps which 1 denoted normal leaf color.

This experiment was conducted in an area where chlorosis was known to have occurred in previous years. It was also known that the occurrence of

¹Journal Paper No. J-14600 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa Project No. 0101.

chlorosis was highly variable and could change greatly in intensity even over areas as small as needed for one block of the experiment.

This paper will use an analysis of covariance to account for spatial variability in the data. Four different covariates will be used, one of which is the Papadakis covariate described by Bartlett (1938). The outcome of the analysis will be described, including the estimates of the regression coefficients of the covariates. Moreover, expected values of the regression coefficients will be derived and compared with the estimates.

2. Demonstration of Spatial Variability

Information on the spatial variability in any given block comes from the residuals of the randomized block analysis. To understand the variability, a display of the residual values in a 31 by 20 array could be made for each block, and the array could be inspected for evidence of spatial trends or contours. Understanding the information in such an array of numbers, however, can be difficult. Therefore, the arrays were simplified by presenting only the signs of the residuals. The arrays of residual signs for blocks 1 and 2 of the experiment are shown in Figures 1 and 2, respectively. Contours were drawn on the displays to enclose the positive residuals (+'s) to aid interpretation. Contours were not drawn around a single + surrounded entirely by -'s or visa versa.

Both Figures 1 and 2 seem to confirm the presence of spatial contours. In Figure 1, for example, the residuals are predominantly negative in the upper right and lower left of the array. Figure 2 shows a large area of negative residuals in the middle of the second block. The use of these displays as evidence of spatial contours may be persuasive to some but not to others. Ultimately, the figures can only be suggestive of the effectiveness of an analysis designed to remove possible spatial variability.

3. Nearest Neighbor Covariates

An analysis of covariance would be successful in accounting for spatial variability if the covariate is highly correlated with the levels of the spatial contours. To determine the contour level for any given plot, information from neighboring plots is used. Thus, when neighboring plots have residuals that are below average, it is assumed that the plot itself will have a residual that is below average. This relation suggests that an average of neighboring residuals would be a measure of the level of the spatial contour for that plot. Based on this reasoning, two different covariates were used in the analysis of the data from the chlorosis experiment. These covariates use the average of 4 and of 14 neighboring residuals, respectively. Figures 3 and 4 show the positions of the neighboring plots (n) in relation to the plot for which the covariate is desired (c). Because plots on the edge or corner of an array do not have neighbors on all sides, the neighbors used to determine the covariate cannot be in the same positions for all plots. The covariate described as using 4 neighboring residuals actually uses only 3 and 2 neighbors for edge and corner plots, respectively (see Figure 3). For the covariate based on 14 neighbors, a slightly different pattern was adopted for edge and corner plots (see Figure 4); this pattern resulted in 14 values being averaged regardless of the position of the plot. The covariate based on

4 residuals corresponds to the Papadakis covariate. The motivation for using 14 residuals is that the covariate has smaller measurement error because it is an average of 14 values rather than 4 values. On the other hand, when the contours change rapidly over short distances, the average of 4 residuals may be able to represent the contours better than an average of 14 residuals because it represents a much smaller area than the average of 14.

An alternative method of calculating the covariates is to use the scores (or y-values) instead of the residuals, because residuals can be difficult to calculate for experiments with a large number of treatments and it would be useful to compare their performance with covariates based on residuals. Thus two additional covariates were used in the analysis of the chlorosis data. These covariates use the average of 4 and 14 neighboring scores, respectively. Notice that use of scores and residuals should be equivalent in describing the spatial contours except that the use of scores should result in covariates with larger measurement errors. This difference in measurement errors implies that the covariates based on scores would be less effective than those based on residuals.

4. Analysis of Covariance Results

Tables 1 and 2 give summary statistics for the analysis of the chlorosis data. Table 1 shows the analysis of variance based on the randomized block model, but without a covariate. The error mean square from the analysis of covariance and the estimate of the regression coefficient for each of the four covariates are shown in Table 2. Also, included is the error mean square from the randomized block analysis (denoted as no covariate). The use of the covariates result in error mean squares that are from 15% to 28% smaller than the value from the randomized block analysis. These are not exceptionally large reductions, but they are highly significant. As expected, the use of residuals resulted in smaller error mean squares than did use of scores (compare .31 with .34 and .29 with .31). The differences, however, are small and not of any practical importance. Moreover, 14 neighbors reduced the error mean square more than 4 neighbors (compare .31 with .34 and .29 with .31). These differences are also small and unimportant. It is evident from Table 2 that the error mean square from the analysis of covariance becomes smaller as the value of the regression coefficient increases.

5. Interpretation of the Regression Coefficients

Insight on how the analysis of covariance should be interpreted can be gained by determining the expected values of the regression coefficients of the covariates. The usual interpretation of a regression coefficient is that it measures the strength of the relation between two variables and that a large coefficient indicates a strong relation. By analogy, this reasoning would imply that a large coefficient for a covariate indicates that much spatial variability is present in the data. But, this interpretation is not correct because the value of the coefficient is related only to the measurement error of the covariate.

The method for deriving the expected value of the regression coefficient differs slightly for each of the 4 covariate types. The covariate calculated as the average of 4 neighboring residuals will be considered first. Consider the 4 neighboring residuals as data from a response

surface where the surface describes the spatial variability in the experiment. Furthermore, it is assumed that the contours of the response surface can be locally approximated by a plane. If least squares is used to fit a plane to the 4 residuals, then the predicted value at the center position is equal to the average of the 4 residuals, and thus it is equal to the value of the covariate. Thus, the covariate is a prediction for the value of the residual at the position in the center of the 4 neighboring residuals.

The fact that the covariate is an estimate of the residual is relevant to the interpretation of the regression coefficient. This relation can be shown by considering the analysis of covariance model:

$$Y_{ij} = \mu + B_i + V_j + \beta X_{ij} + e_{ij}, \quad (1)$$

where μ , B_i , V_j , X_{ij} , e_{ij} , and β are the overall mean, block effect, variety effect, covariate, error from the analysis of covariance, and regression coefficient, respectively. This expression can be rearranged in the form

$$Y_{ij} - \mu - B_i - V_j = \beta X_{ij} + e_{ij} \quad (2)$$

The terms at the left of the equal sign define the true value of the residual from the randomized block analysis, whereas the covariate X_{ij} is an estimate of the same value. Thus equation (2) represents a simple linear regression in which the dependent and independent variables represent the same quantity and the regression coefficient equals 1.0. The estimate of β from the analysis of covariance, however, will be biased because the covariate is measured with error. In general, the expected value of the regression coefficient from a simple linear regression is

$$E(\hat{\beta}) = K\beta, \quad (3)$$

where $K = \sigma_X^2 / (\sigma_X^2 + \sigma_U^2)$ and σ_X^2 and σ_U^2 are the variances of the independent variable X and the measurement error of X , respectively (Snedecor and Cochran, 1989, p. 173). This formula applies to the analysis of covariance because β equals 1.0, σ_X^2 is the variance of the true residuals from the randomized block analysis, and σ_U^2 is the measurement error of the covariate. The error mean square from the randomized block analysis, s^2 , can be used to estimate σ_X^2 . The variance of the measurement error of X can be estimated by $s^2/4$ because the covariate is the average of 4 residuals, each of which has an estimated variance of s^2 . Thus the expected value of the regression coefficient is approximately estimated by

$$s^2 / (s^2 + s^2/4) = .8 \quad (4)$$

The numerical value of s^2 is not important in determining the expected value because it can be cancelled in numerator and denominator. This expected value is only an approximation because the residuals are not exactly independent, and hence the variance of the measurement error is

not estimated by $s^2/4$. The approximation, however, is useful because it shows that the value of the regression coefficient is unrelated to the strength of the spatial variability in the data. Rather, the value of the regression coefficient is entirely determined by the variance of the measurement error of the covariate. Interestingly, the estimated value of the regression coefficient from the analysis of covariance (see Table 2) agrees exactly with the expected value.

Similar logic can be used to determine the expected value of the regression coefficient when the average of 4 neighboring scores is used as a covariate. Recall that the justification for use of the average of neighboring scores as a covariate (see section 3) was that the scores should estimate the same spatial trends as the residuals. The variance of the scores, however, will be larger than that of the residuals because of the extra variability of the varieties and blocks that will be present in the scores. The presence of the variety influences are not expected to cause biased estimates of the spatial contours because the varieties were randomly assigned within blocks. The variance of the measurement error of the covariate would be $(\sigma_b^2 + \sigma_v^2 + \sigma_e^2)/4$, where σ_b^2 , σ_v^2 , and σ_e^2 are the variance components of the blocks, varieties, and residuals, respectively. Estimates of these components were calculated by equating the mean squares in the randomized block analysis with their expected values and solving the resulting equations. These estimates are $s_b^2 = .03$, $s_v^2 = .51$, and $s_e^2 = .40$. The estimates of σ_x^2 and σ_u^2 are $.03 + .51 + .40$ or $.94$ and $.94/4$ or $.24$, respectively. When these estimates are used in equation (3), the value of the regression coefficient is $.63$. This value is greater than the estimated coefficient ($.39$) from the analysis of covariance.

The expected values of the covariate regression coefficients based on 14 neighbors differ from those based on 4 neighbors because the variance of the measurement error differs. That is, the covariate is the average of 14 values rather than 4 values. In this instance, σ_u^2 is estimated by $s^2/14$ and $.94/14$, and the expected values are $.93$ and $.86$ for 14 residuals and 14 scores, respectively.

Table 3 shows the values of the regression coefficients for each of the 4 covariate types when estimated from the analysis of covariance and when variance component estimates are used in equation (4) to obtain expected values. The agreement between the estimated and expected coefficients is closest for the covariate based on 4 residuals. The lack of agreement for the two covariates based on 4 and 14 scores could be due to inaccurate estimates of the variance of the measurement error of the covariate and/or the variance of the true residuals. The difference between the estimated and the expected coefficients for 14 residuals cannot be explained by inaccurate variance estimates because the variance terms cancel in the formula for the expected value. Furthermore, the expected value is smaller than 1.0, but its estimated value is significantly greater than 1.0. A possible explanation for this discrepancy is that use of 14 residuals gives biased estimates of the residual at the center because of non-linearity in the spatial variation. Use of a biased covariate could result in a coefficient greater than 1.0.

6. Summary

Of the 4 covariates used to account for spatial variability in the experiment, each reduced the error mean square significantly when compared with the randomized block analysis without a covariate. Covariates calculated from residuals reduced the error mean square more than those based on scores. Moreover, the use of 14 neighbors reduced the error more than 4 neighbors. The differences among the covariates, however, were small and not of practical importance.

The expected values of the regression coefficients were related to the measurement error of the covariate and not to the strength of the spatial variability in the data. The result is that their values range from 0 to 1. Except for the covariate based on 14 residuals, all coefficients estimated from the analysis of covariance were in this range. This exception is noteworthy because it casts doubt on the validity of interpretations of the adjusted treatment means from the analysis of covariance.

Bartlett (1978) used a two-dimensional generalization of a Markovian model to derive the expected value of the regression coefficient for the covariate based on 4 residuals. The expected value for this model is a function of the model correlation coefficients and can be as large as unity. In comparison, the expected value determined from the measurement error model does not depend on any parameters and equals 0.8, which is less than unity.

References

- Barlett, M.S. 1938. The approximate recovery of information from field experiments with large blocks. *J. Agric. Sci.* 28:418-427.
- Bartlett, M.S. 1978. Nearest neighbour models in the analysis of field experiments. *J.R. Statist. Soc. B* 40:147-174.
- Snedecor, G.W. and Cochran, W.G. 1989. *Statistical Methods*, Eighth Edition, Ames, Iowa: Iowa State University Press.

Table 1. Analysis of variance of chlorosis scores of 620 soybean varieties from a randomized block design

Source	d.f.	M.S.
Blocks	3	16.73
Varieties	619	2.45
Error	1857	.40

Table 2. Residual mean square and covariate regression coefficient from the analysis of covariance for 4 covariate types and residual mean square from the randomized block analysis

Covariate Calculated From the Average of:	Residual M.S.	Regression Coefficient
4 scores	.34	.39
14 scores	.31	.76
4 residuals	.31	.80
14 residuals	.29	1.16
randomized block analysis with no covariate	.40	--

Table 3. Regression coefficients for 4 covariate types estimated from the analysis of covariance and their estimated expected values

Covariate Calculated From the Average of:	Covariate Regression Coefficient	
	Estimated From the Analysis of Covariance	Estimated Expected Value
4 scores	.39	.63
14 scores	.76	.86
4 residuals	.80	.80
14 residuals	1.16	.93

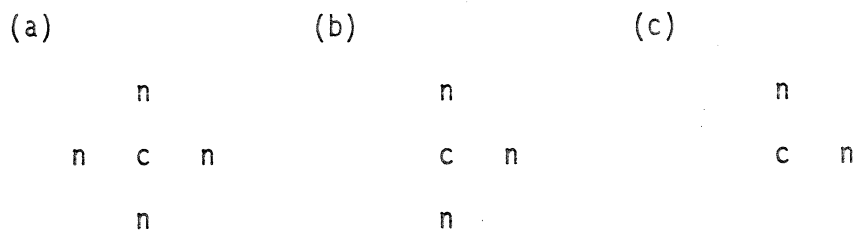


Figure 3. Position locations, n, of neighboring plots used for calculation of a 4-neighbor covariate for position c. Neighbors used when the plot is (a) interior in the block, (b) at the block edge, and (c) at the block corner.

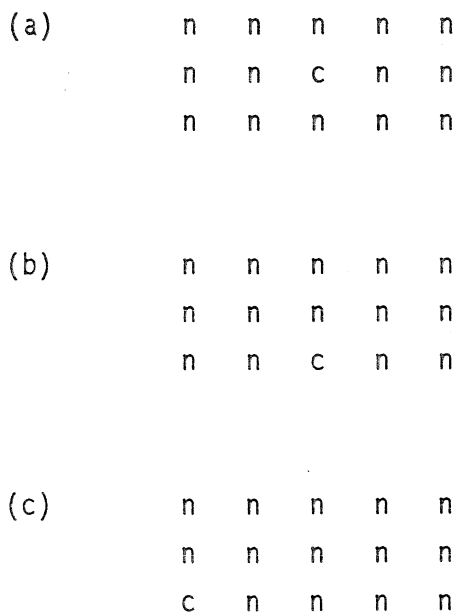


Figure 4. Position locations, n, of neighboring plots used for calculation of a 14-neighbor covariate for position c. Neighbors used when the plot is (a) interior in the block, (b) at the block edge, and (c) at the block corner.

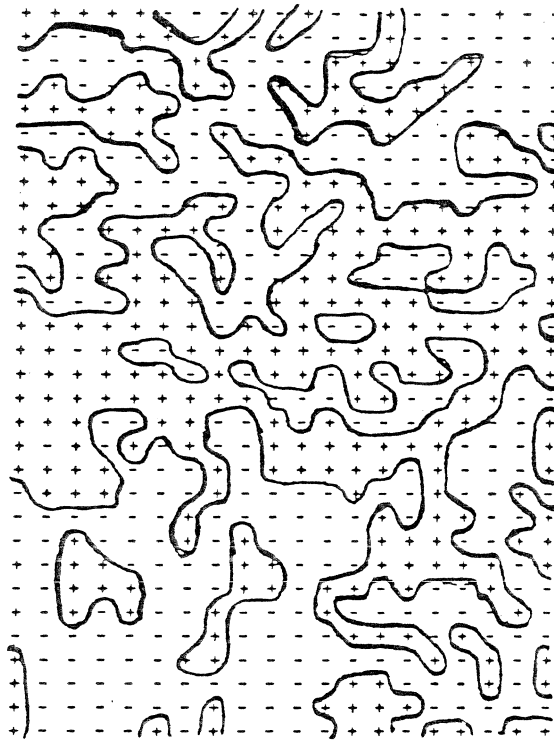


Figure 1. Residual signs for block 1 displayed in their original field positions.

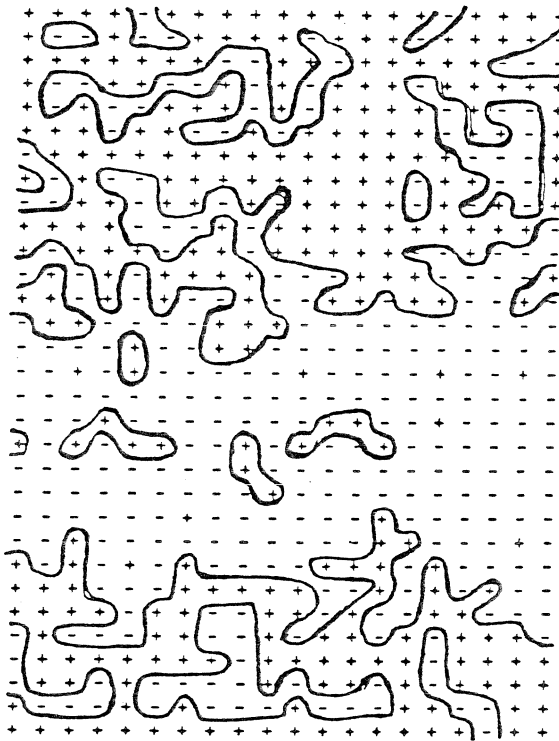


Figure 2. Residual signs for block 2 displayed in their original field positions.