

Combining Standard Block Analyses With Spatial Analyses Under a Random Effects Model

Walter T. Federer, Elizabeth A. Newton, and Naomi S. Altman*

*Cornell University
United States*

Abstract

Spatial trends are often a significant source of variability in field trials. Since trends vary from block to block, they should be estimated as random effects. In this paper we propose to consider spatial covariates as post hoc random effects within the context of the experiment design. We demonstrate that making use of the spatial information leads to more efficient estimation of treatment effects. The models considered for spatial effects include block effects which are part of the experiment design, random gradients, regression trends, nearest neighbor analysis and smoothing. The analyses are applied to an example, which exhibits quite different results for the different methods. Since computations are tedious, programs for the various statistical procedures are presented.

Key words and phrases: ANOVA, gradients, field trials, trend analysis, nearest neighbor, random effects, REML, smoothing, spatial analysis.

1 Introduction

Spatial trend is often a significant source of variation among experimental units in field trials. We propose a number of methods for accounting for smooth spatial variation which varies randomly from block to block. We demonstrate that accounting for spatial variation can improve the efficiency of estimating treatment effects, and may improve on removing row and column effects, even when the experiment has been designed to account for such effects. As well, we provide sample SAS code to obtain the maximum likelihood estimators (MLEs) of the treatment effects in the context of the linear mixed model.

*The authors can be reached at the Biometrics Unit, 434 Warren Hall, Cornell University, Ithaca, NY 14853 or by e-mail at biometrics@cornell.edu.

A number of features of a standard analysis may alert the scientist to the need to consider spatial variation. The most effective diagnostic is often mapping the residuals on the field design. Patterns of positive and negative residuals may indicate spatial gradients. When the data are counts or indicators, a mean square error that is too large, indicating extra-Poisson or extra-Binomial variation, may also indicate the need to remove spatial trends. For some responses, the order of magnitude of the coefficient of variation is well known. For example, depending on plot size, cereal yield experiments usually have a coefficient of variation of 4–8% and maize of 8–12%. Larger coefficients of variation may also indicate an uncontrolled source of variation. See Federer (1992) for other diagnostics.

In the next section, a number of spatial analyses are introduced in the context of a lattice square experiment (Table 12.5 of Cochran and Cox, 1957). The models presented will be for this lattice square experiment, but it is straightforward to extend the procedures to other designs and situations. The key ideas are that spatial gradients should be smooth, and that the gradients vary randomly among blocks. A designed experiment should be analyzed with spatial effects taken into account, while retaining any restrictions to randomization that were part of the design. In addition, if the experimental design is not balanced with respect to the spatial gradients, treating the gradient as a random effect may substantially alter the MLEs of the treatment effects compared to their ANOVA estimators. The MLEs may be considered the “adjusted treatment means” in analogy with the adjusted treatment means obtained from classical incomplete block analyses.

1.1 The Data

The numerical example used for the analyses was presented by Wadley (1946) and is reproduced in Table 12.5 of Cochran and Cox (1957). The data are means of counts of three samples of 100 cotton squares indicating attack or not of boll weevils. The mean count for the experiment was 11. If the counts were distributed as a Poisson variable, one would expect the residual mean square to be near $11/3 = 3.667$, since the counts are means of three samples. The anticipated coefficient of variation would be $1.91/11 = 17\%$. The experiment design is a balanced lattice square with $v = 16$ insecticide treatments arranged in $r = 4$ rows and $c = 4$ columns within each of $b = 5$ complete blocks or replicates. The randomization is restricted in such a manner as to have every ordered pair of treatments appear together once in each row and once in each column. Restrictions on randomization of treatments must be taken into account when an analysis is made. SAS code for each analysis is presented in the appendix.

2 Models for Analysis of Field Trials with Spatial Gradients

A number of models can be devised for taking into account spatial effects in field trials. In this section, we present a few possibilities. In the next section, the results of these models are discussed for the example.

2.1 Standard Textbook (ANOVA) Model

The first form of a spatial analysis is the usual textbook response model for a lattice rectangle designed experiment (See e.g., Cochran and Cox, 1957, Federer, 1955, and Kempthorne, 1952.):

$$Y_{ghij} = \mu + \beta_g + \rho_{gh} + \gamma_{gi} + \tau_j + \epsilon_{ghij} \quad (1)$$

where μ is a general mean effect, β_g is the g^{th} replicate effect distributed with mean zero and variance σ_β^2 , ρ_{gh} is the gh^{th} row effect distributed with mean zero and variance σ_ρ^2 , γ_{gi} is the gi^{th} column effect distributed with mean zero and variance σ_γ^2 , τ_j is the j^{th} treatment effect, ϵ_{ghij} is a random error effect distributed with mean zero and variance σ_ϵ^2 , $g = 1, \dots, b$, $h = 1, \dots, r$, $i = 1, \dots, c$, and $j = 1, \dots, v = rc$.

The standard intrablock effects analysis of variance (ANOVA) can be obtained by using, for example, SAS PROC GLM, treating replicate, row(replicate) and column(replicate) as fixed. To compute the MLEs of the treatment effects using REML (restricted maximum likelihood) solutions for the variance components, SAS PROC MIXED could be used. Sample SAS code is displayed in the appendix (5.1).

Sums of squares and intrablock means obtained are those presented in textbooks. Textbook analyses make use of ANOVA solutions for variance components when recovering interblock information, whereas the REML solutions automatically adjust for interblock information. The adjusted treatment means can vary considerably between the methods for some situations since the estimated variance components for REML and ANOVA can be quite different.

2.2 Differential Trends within Blocks (Rows)

Cox (1958) considered a situation wherein differential curvatures existed within the columns of a Latin square. Differential trends may also occur within each incomplete block of an incomplete block design or within each row (column) of a lattice rectangle (square) experiment (Federer, 1996). In such an event, equation (1) is an inappropriate response model. Instead, the following response model is used:

$$Y_{ghij} = \mu + \beta_g + \rho_{gh} + \pi_{gh}\alpha_{ghi} + \tau_j + \epsilon_{ghij} \quad (2)$$

where the column effects have been replaced by a linear trend within each row. In this model π_{gh} is the gh^{th} linear regression coefficient of responses on the

ordered and centered positions, a_{ghi} , within a block and the remaining terms are as defined for equation (1). The linear regression coefficients are random variates distributed with mean zero and variance σ_π^2 . (See Federer, 1996, for other models.) If desired, additional polynomial regressions could be added to the model, depending upon the nature of the trends within blocks or rows (columns).

A continuous variable, a_{ghi} , for position in the row (column) needs to be added to the data. (For example, the coefficients of the corresponding linear contrast can be used.) The standard analysis using SAS PROC GLM and most other analysis of covariance (ANCOVA) software treats the regression coefficients as fixed effects. To compute the adjusted treatment means (MLEs) using the REML solutions for the random slopes, PROC MIXED or other general linear mixed model software must be used. Sample SAS code is displayed in the appendix (5.2). The code is readily generalized to polynomial trends.

2.3 Row and Column Regressions

The analysis above is readily generalized to the case of polynomials in both row and column. In this case, the spatial trend need not be aligned to the row and column design of the experiment. (This situation has been considered by several authors over the past 50 years or more; also, see Federer, 1996, for the random effects situation.). A response model for this type of variation is:

$$Y_{ghij} = \mu + \beta_g + \tau_j + \pi_{g1}RL_{ghi} + \pi_{g2}RQ_{ghi} + \pi_{g3}CL_{ghi} + \pi_{g4}CQ_{ghi} + \pi_{g5}LL_{ghi} + \pi_{g6}LQ_{ghi} + \pi_{g7}QQ_{ghi} + \epsilon_{ghij} \quad (3)$$

where RL_{ghi} are the linear regression values of ordered row positions, RQ_{ghi} are the quadratic regression values of ordered row positions, CL_{ghi} and CQ_{ghi} are defined similarly, LL_{ghi} is the row linear by column linear interaction, LQ_{ghi} is the row linear by column quadratic interaction, QQ_{ghi} is the row quadratic by column quadratic interaction, and the remaining terms are as defined for (1). As in (2), the regression coefficients, π_{gk} are considered to be random effects. If appropriate, additional regression terms may be included or terms may be deleted from the model.

Once again, standard ANCOVA software can be used to remove such trends if they are considered fixed. More realistically, however, the trends should be considered random effects and may be estimated by use of PROC MIXED or other linear mixed model software. An example is in the appendix (5.4).

Note that other higher degree regression polynomials or other forms of regression could be used as well. This analysis preserves the design structure. The above analysis is a form of what has been called trend analysis in the spatial statistics literature. It is, however, a technique that goes back at least to R. A. Fisher over 60 years ago for the fixed effects case.

2.4 Nearest Neighbor Analysis

A response model equation for a nearest neighbor analysis (Papadakis, 1937) of a design laid out in a row-column arrangement is:

$$Y_{ghij} = \mu + \beta_g + \tau_j + \pi_{g1}R_{ghi} + \pi_{g2}C_{ghi} + \pi_{g3}P_{ghi} + \epsilon_{ghij} \quad (4)$$

where $R_{ghij} = (\epsilon_{g(h-1)ij} + \epsilon_{g(h+1)ijk})$, $C_{ghij} = (\epsilon_{gh(i-1)j} + \epsilon_{gh(i+1)j})$ are the averages of adjacent row and column errors respectively, and P_{ghij} is the interaction (product) of these terms, π_{gk} are regression coefficients and the other terms are as defined in (1). This model induces a spatial covariance structure on the errors. Papadakis (1937) used a two-stage procedure in which the errors were estimated by the residuals from a RCB analysis, and the π_{gk} were estimated as fixed effects by using these residuals in an ANCOVA. The MLEs of all coefficients and treatment effects under a fixed effects model can be computed by iteratively fitting model (4), using the residuals from the previous iteration to estimate R_{ghi} , C_{ghi} and P_{ghi} (Papadakis, 1937). Alternatively, REML can be used to fit the covariance structure and obtain the MLEs of the treatment effects. An interesting discussion of this is in Cressie (1993, Chap. 5.7).

However, for random blocks the π_{gk} should be random effects. Discussion of fitting a covariance model with random parameters for the covariance structure is in Section 2.6. As a “quick and dirty” approximation, we used Papadakis’ (1937) two-step procedure, using SAS PROC GLM to generate the RCB residuals and PROC MIXED to estimate the regression coefficients as random effects. The program is listed in the appendix (5.4).

2.5 Smoothing

Polynomial trends (equations 2 or 3) can be replaced by general smooth trends. If the trends are assumed to line up with the column (row) layout of the design, the polynomial trends in (2) can be replaced with the model:

$$Y_{ghi} = \mu + \beta_g + \rho_{gh} + \gamma_{ghi} + \tau_j + \epsilon_{ghij} \quad (5)$$

where γ_{ghi} is a smooth trend within row ρ_{gh} . If the trends are not assumed to line up with the rectangular layout of the design, then polynomial trends of the type in model (3) can be replaced with the model:

$$Y_{ghij} = \mu + \beta_g + \gamma_{ghi} + \tau_j + \epsilon_{ghij} \quad (6)$$

where γ_{ghi} now represents a 2-dimensional spatial curve. In either case, we now have a semi-parametric additive model, (Hastie and Tibshirani, 1990), sometimes termed a partial linear model (Heckman, 1986; Speckman, 1988), and a number of methods have been developed for simultaneously fitting the smooth and parametric terms in the fixed effect case. The smooth terms in the model are fitted using a tuning parameter (the bandwidth, span or smoothing parameter) which controls the smoothness of the spatial term. For the fixed effects case, the tuning parameter should be fitted adaptively in each block, as there

is no a priori reason to expect the spatial trends to have the same degree of smoothness in different blocks.

Smoothing splines (Wahba, 1990) and least squares smoothing (Green, Jennison and Seheult, 1985; Jennison and Seheult, 1984) have natural extensions to the random effects case, as discussed in the papers cited. In this paper, we have followed the computational procedure outlined by Green et al (1985), using S-PLUS as the computational tool, to fit (5). A copy of the S-PLUS code is available upon request.

2.6 Covariance Models

Models (6) and (7) treat spatial gradients within a block as fixed in the sense that the gradient should persist under a new realization of the errors. Often spatial gradients are modeled as random correlated processes. A large literature exists for this approach in the spatial and geostatistics literature. See Cressie (1993) for a comprehensive treatment. The model is:

$$Y_{ghi} = \mu + \beta_g + \gamma_{ghi} + \tau_j + \epsilon_{ghij} \quad (7)$$

where all the terms are as in (5.2) except that γ_{ghi} is now the random realization of a spatially autocorrelated process $\gamma_{ghi} \sim (0, \Sigma(\theta_g))$ where θ_g are parameters defining the correlation structure. If the θ_g are considered fixed, this model may be fitted using REML to estimate the covariance structure. For θ_g random we plan to use Monte Carlo Markov Chain methods for fitting hierarchical models (Geyer, 1992). We have not yet explored the intricacies of fitting such a model.

3 Data Analysis

In this section we apply the methods above to the lattice square experiment described in Table 12.5 of Cochran and Cox (1957). As well, we present a randomized complete block design analysis which, however, is inappropriate as it ignores the restrictions to randomization imposed by the design and row and column effects for which the design should adjust.

In Table 1, Type III (partial) sums of squares are presented for treatments, spatial effects, and residual from each model. (The block mean has been removed from the data prior to analysis, accounting for 4 degrees of freedom in the model.) The spatial sum of squares is an average of all the blocking variables within replicates. The F-value is the ratio of treatment and residual mean squares.

In the RCB analysis, there appear to be significant treatment effects. However, a plot of the residuals within block shows that there is considerable within-block spatial pattern which does not align with the row/column layout of the treatments. As well, the residuals are quite skewed, possibly indicating confounding of treatment and residual effects. Finally, the residual mean square is much greater than we would expect from Poisson data.

The standard lattice square analysis, LSD(1), suggests that variation among treatments is less than the residual mean square. The residuals are somewhat smaller than those obtained from the RCB analysis, but the treatment mean square has also been considerably reduced. Spatial patterning of the residuals is also evident, but is less clear than in the RCB analysis.

The remaining analyses treat the spatial effects as continuous covariates with random coefficients. It is interesting to note that the treatment and residual mean squares differ considerably depending on the model.

RG(2), CG, CLGr and RCreg(3) are all models in which the spatial effects are considered to be polynomials in rows and columns with random coefficients. RG(2) is model (2), with a separate linear gradient in each column. CG has a separate linear gradient in each column. CLGr assumes a linear gradient across the columns in each block, and a linear gradient within each column. RCreg(3) is model 3. It is interesting to note that the F-ratio varies considerably among these models, as does the amount of variation ascribed to the treatment and spatial effects. However, the designs in which the gradient is assumed to be aligned with the row/column layout of the design (LSD(1), RG(2), CG and CLGr) all have about the same partial sum of squares for spatial effects, although they have different degrees of freedom. Conclusions drawn from these models range from “no significant treatment effects” (RG and CG) to “highly significant treatment effects” (RCreg(3)).

Of the linear models, model (3) appears to give the best fit, with a residual mean square of 11.9. However, this is still three times the variance which would be indicated if the counts were Poisson.

If the experimenter had used the standard textbook analysis, he would reach the conclusion that there were no differences among the 16 insecticide treatments. Using the RCreg(3) analysis, it would be concluded that there were significant differences at the 2% level. These two completely different conclusions demonstrate how spatial patterns can sometimes distort treatment differences and why it is essential that an appropriate analysis be selected.

4 Discussion

For any given experiment design, a number of analyses are generally possible, besides the standard textbook one. Although the experiment design should, to the extent possible, control for known sources of variation, this is often not possible due to the inherent variability of the experimental material. When the experiment has a spatial component, this is particularly important, as field plots may have to adapt to existing field designs which have not been laid out taking spatial variability into account. Even when the field has been optimally oriented with respect to existing gradients, the gradients may not be orthogonal, so that row/column designs cannot fully control for spatial effects.

It is interesting to note the results presented by Kempton (1984) and Kempton and Howes (1981) on the relative efficiencies of standard textbook lattice square analyses and some spatial methods of analysis. They studied the results

Method of analysis	mean	median	range
Rows as incomplete blocks	124	112	91-270
Lattice square	147	136	95-312
1D row neighbors	138	124	96-282
column neighbors	117	108	98-211
maximum of row or column neighbor	144	133	98-282
Papadakis (2 covariates)	148	136	96-297

of 118 wheat variety trials all designed as five by five balanced lattice squares. See the results in the figure above.

In these trials, the experimental unit was three times as long as wide so that the distance between rows was three times larger than between columns. So analyses involving columns will not control as much variation as ones with rows. An interesting statistic not reported would be the number of experiments in which each of the above analyses had the highest efficiency. The table demonstrates clearly the need to remove row and column effects, although for these trials the lattice square analysis appears to be just as effective as the more complex spatial models.

A possible contender for the analysis R_Creg(3) using response model equation (3) would be to replace LL, LQ, and QQ with first, second, and perhaps third principal components in the manner used for additive main effects and multiplicative interactions (AMMI) analysis (Gaugh, 1988). The interactions used in (3) may not maximize the sums of squares for variation whereas the principal components method would do this. Since this method is not for interpretive purposes but to control extraneous variation, there would be no problem in using an AMMI analysis.

As we have seen in our preceding numerical example as well as the example above, a number of a priori reasonable models can lead to very different interpretation of treatment effects. There is little guidance in the literature for choosing among non-nested models of the types presented here, beyond graphical analysis of the residuals. Serious concerns when choosing among several models are overfitting and bias. We are studying these via simulation studies. For the semiparametric methods such as nearest neighbors, smoothing and AMMI the appropriate degree of freedom adjustment to the error mean square also needs to be computed.

References

- [1] Cochran, W. G., Cox, G. M. (1957). *Experimental Designs*, 2nd edition. Wiley & Sons, Inc. New York.
- [2] Cox, C. P. (1958). The analysis of Latin squares with individual curvatures in one direction. *J. Royal Statistical Soc., Series B*, 20, 193-204.

- [3] Cressie, N.A.C. (1993) *Statistics for Spatial Data*. John Wiley & Sons: New York.
- [4] Federer, W. T. (1992) Diagnostic Procedures for Analysis of Variance. In *New Progress in Probability and Statistics, Proceedings of the Conference "Nicolas Arriquibar", 4th International Meeting of Statistics in the Basque Country*. M.L. Puri and J.P. Vilaplana (eds.) International Science Publishers, Zeist, The Netherlands, pp 1-23.
- [5] Federer, W. T. (1955). *Experimental Design - Theory and Application*. Macmillan Co., New York. (Republished by Oxford & IBH Publishing Co., Calcutta, 1967 and 1974).
- [6] Federer, W. T. (1996). Recovery of interblock, intergradient, and intervariety information for incomplete block and lattice rectangle designed experiments. BU-1315-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY 14853.
- [7] Federer, W. T., Wolfinger, R. D. (1996). SAS PROC GLM and PROC MIXED for recovering inter-effect information. BU-1330-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY 14853.
- [8] Gauch, H. G. Jr. (1988) Model selection and validation for yield trials with interaction. *Biometrics*, **44**, 705-715.
- [9] Geyer, C.J. (1992) Practical Markov Chain Monte Carlo (with discussion). *Statistical Science*, **7**, 473-502.
- [10] Green, P., Jennison, C., Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *J. Royal Statistical Soc., Series B*, **47**, 299-315.
- [11] Hastie, T.J., Tibshirani, R.J. (1990) *Generalized Additive Models*. Chapman & Hall: London.
- [12] Heckman, N.E. (1986) Spline smoothing in a partly linear model. *J. Royal Statistical Soc., Series B*, **48**, 244-248.
- [13] Kempthorne, O. (1952) *The Design and Analysis of Experiments*. John Wiley & Sons, Inc., New York.
- [14] Kempton, R. A. (1984). Comparison of nearest neighbor and classical methods of analysis. In *Spatial Methods in Field Experiments* (Editor: R. A. Kempton), University of Durham, pp 51-52.
- [15] Kempton, R. A., Howes, C. W. (1981). The use of neighboring plot values in the analysis of plot trials. *Applied Statistics*, **30**, 59-70.
- [16] Jennison, C., Seheult, A. (1984). Two dimensional (2D) least squares smoothing (LSS) analysis. In *Spatial Methods in Field Experiments* (Editor: R. A. Kempton), University of Durham, pp. 25-28.

- [17] Papadakis, J. S. (1937). Methode statistique pour des Experiences sur Champ. *Bulletin Inst. Ameliorations Plantes Salonique*, **23**, 1-30.
- [18] Speckman, P. (1988) Kernel smoothing in partial linear models. *J. Royal Statistical Soc., Series B*, **50**, 413-436.
- [19] Stroup, W. W., Miltz, D. K. (1991). Nearest neighbor adjusted best linear unbiased predictor. *American Statistician*, **45**, 194-200.
- [20] Wadley, F. M. (1946). Incomplete block designs in insect population problems. *J. Economic Entomology*, **38**, 651-654.

Source of variation	df	Sum of squares	df	Sum of squares	df	Sum of squares
	RCBD		LSD(1)		RG(2)	
Treatment	15	1244	15	320	15	347
Spatial	0	0	30	1653	35	1650
Residual	60	2333	30	680	25	474
F-ratio		2.13		0.94		1.22
	CG		CLGr		RCreg(3)	
Treatment	15	514	15	614	15	435
Spatial	35	1645	25	1637	30	1761
Residual	25	614	35	695	30	357
F-ratio		1.39		2.06		2.43
	NN(4)		SMOOTH(5.2)			
Treatment	15	530	15	723		
Spatial	?	920	27.1	2147		
Residual	?	1151	36.9	738		
F-ratio		?		2.41		

Table 1: Partial sums of squares for treatment, spatial, and residual for various analyses// RCBD = randomized complete block; LSD(1) = lattice square, equation (1); RGr(2) = rows and gradients in rows, equation (2); CGr = columns and gradients in columns; CLGr = linear column effects within replicates with differential gradients in columns; RCreg(3) = linear and quadratic regressions and interactions as in equation (3); NN(4) = nearest neighbor within replicates as in equation (4); SMOOTH(5.2)=Least squares smoothing as in equation 5.2 ; F-ratio = treatment / residual.

5 Appendix - SAS Programs

5.1 Lattice Square (Model 1)

ANOVA

```
data lsgr;
  infile 'lsgr1645.dat';
  input count rep row column treat;
proc glm data = lsgr;
  class rep row column treat;
  model count = rep row(rep) column(rep) treat;
  random rep row(rep) column(rep);
  lsmeans treat;
run;
```

Using REML to Obtain MLE of Treatment Effects

```
data lsgr;
  infile 'lsgr1645.dat';
  input count rep row column treat;
proc mixed data = lsgr;
  class rep row column treat;
  model count = treat;
  random rep row(rep) column(rep);
  lsmeans treat;
run;
```

5.2 Random Gradients within Row (Model 2)

The created variable “grad” is the coefficients of the linear contrast of columns, within each row.

ANOVA

```
data lsgr;
  infile 'lsgr1645.dat';
  input count rep row column grad treat;
proc glm data = lsgr;
  class rep row column treat;
  model count = rep row(rep) grad*row(rep) treat;
  random rep row(rep);
  lsmeans treat;
run;
```

Using REML to Obtain MLE of Treatment Effects

```
data lsgr;
  infile 'lsgr1645.dat';
  input count rep row column grad treat;
proc mixed data = lsgr;
  class rep row treat;
  model count = treat;
  random rep row(rep) grad*row(rep);
  lsmeans treat;
run;
```

5.3 Row and Column Regressions (Model 3)

RL and RQ are the linear and quadratic within row variables. CL and CQ are the linear and quadratic within column variables.

ANOVA

```
data lsgr;
```

```

infile 'lsgr1645.dat';
  input count rep treat RL RQ CL CQ;
  LL = RL*CL;
  LQ = RL*CQ;
  QQ = RQ*CQ;

proc glm data = lsgr;

class rep treat;
model count = rep RL*rep RQ*rep CL*rep LL*rep LQ*rep QQ*rep treat;
random rep;
lsmeans treat;

run;

```

Using REML to Obtain MLE of Treatment Effects

```

data lsgr;

infile 'lsgr1645.dat';
input count treat RL RQ CL CQ;
LL = RL*CL;
LQ = RL*CQ;
QQ = RQ*CQ;

proc mixed data = lsgr;

class rep treat;
model count = treat;
random rep RL*rep RQ*rep CL*rep LL*rep LQ*rep QQ*rep;
lsmeans treat;

run;

```

5.4 Two-step NN Analysis

```

data lsgr;
  infile 'a:\lsgr1645.dat';
input count rep row column treat;

proc sort; by rep row column;
proc glm data = lsgr;

class rep treat;
  model count = rep treat;
  output out = lsgr2 r = res;
  data lsgr3;
  merge lsgr lsgr2;

proc iml;
use lsgr3;

```

*sort data
generate RCB residuals*

*generate R, C and P, taking edge effects
into account*

```

read all var {rep row column res};
n = nrow(res);
rnn = J(n,1,0);
cnn = J(n,1,0);
nr = max(row);
nc = max(column);
nb = max(rep);
ind = 0;
do k = 1 to nb;

do i = 1 to nr;
do j = 1 to nc;
ind = ind + 1;
if i = 1 then rnn[ind] = res[ind + nc];
else if i = nr then rnn[ind] = res[ind - nc];
else rnn[ind] = 0.5*(res[ind - nc] + res[ind + nc]);
if j = 1 then cnn[ind] = res[ind + 1];
else if j = nc then cnn[ind] = res[ind - 1];
else cnn[ind] = 0.5*(res[ind - 1] + res[ind + 1]);
end;
end;

end;
create lsgr4 {rnn cnn};
append;
quit;
data lsgr5;
merge lsgr3 lsgr4;

proc mixed data = lsgr5;                                Compute REML estimators.

class rep treat;
model count = treat;
random rep rnn*rep cnn*rep rnn*cnn*rep;
run;

```