GENSTAT and S-PLUS Codes For Recovering Interblock and Intergradient Information

by
John Barnard, Computer Center, Food Research Bldg., N. Y. Agricultural Experiment Station, Geneva, NY 14456-0462, JB18@cornell.edu
Walter T. Federer, 434 Warren Hall, Cornell University, Ithaca, NY14853, wtf1@cornell.edu

BU-1381-M
January, 1997

Abstract

In order to facilitate the computational needs of researchers, computer programs using the GENSAT and S-PLUS software packages are presented. In particular, codes are presented for both fixed and mixed model analyses for an incomplete block design, for a lattice rectangle design, and for differential gradients within incomplete blocks or within the rows (columns) of a lattice rectangle design. The block or row and column effects are treated as random effects in order obtain a more efficient analysis of experimental data and to make use of the information contained in inter-effect mean squares. Also, differential gradients may occur within incomplete blocks or rows (columns) and need to be treated as random effects. The GENSTAT package is widely used by experimenters, especially outside the United States. The S-PLUS package is gaining in popularity among researchers, especially statisticians.

1. Introduction

When analyzing data from experiments designed as incomplete block or row-column designs, an efficient analysis recovers the treatment information contained in the incomplete blocks or in the rows and columns of the design and uses a mixed model procedure. The blocks and the rows and columns are considered to be random effects., resulting in the so-called mixed model situation. The resulting adjusted means containing this additional information have a smaller mean square error than means ignoring this information.

Textbook analyses may be appropriate for certain types of experimental variation but quite inappropriate for other types. Patterns of spatial variation not anticipated may arise during the course of conducting the experiment. For example, invasion of insects may not occur in the manner used for blocking. Water or freezing damage to experimental units may occur in patterns not anticipated by the experimenter. Differential gradients may occur within the complete blocks or within the incomplete blocks. In such instances, textbook analyses will be inappropriate and need to be replaced by an appropriate analysis. The differential gradients should be considered as random effects. .

This paper demonstrates how to program analyses on GENSTAT and S-PLUS for a triple lattice (incomplete block) designed experiment, for a lattice square (lattice rectangle) designed experiment, and for differential linear gradients within the rows (or incomplete blocks) of a lattice square designed experiment. GENSTAT codes are presented for the ANOVA, regression (fixed effects model), and variance component using REML (restricted maximum likelihood) analyses. Comparable S-PLUS codes are given for the variance component analysis using REML solutions for the variance components. It should be noted that different solutions for the variance components are obtained from the two programs and hence, the adjusted means will differ to some extent. S-PLUS results are the same as obtained by SAS PROC MIXED. It appears that the developers of the packages used different boundary conditions for their REML solutions of variance components.

2. Recovery of Interblock Information in Incomplete Block Designs

The GENSTAT regression program is a fixed effects or intrablock modeling of the responses. The information in the blocks is ignored and intrablock treatment means are obtained. The code for the incomplete block data set named fed933.dat, Example XI-3 in Federer (1955), is:

```
"gn-1-1  Triple lattice"
units [27]
factor r, b, t
open 'fed933.dat'; channel = 2
read [channel = 2] y, r, b, t;
model y
terms r/b + t
fit [print=accumulated; fprobability=yes] r/b + t
predict t
stop
```

The first line names the analysis, the second gives the number of responses to be analyzed, and the third line lists the factors (class variables in SAS notation). The fourth line opens the data set named fed933.dat. The fifth line denotes the input from fed933.dat and the entries must be in the correct order with Y = YIELD or RESPONSE, R = REPLICATE or COMPLETE BLOCK, B = INCOMPLETE BLOCK, and T = TREATMENT. The sixth line states that response Y is being modeled and the seventh line gives the terms of the linear model used *sans* the error term. The eighth line requests a nested or accumulated (Type I) analysis of variance table with probability statements for F-values. The ninth line obtains intrablock (fixed effects) means with their standard errors.

The GENSTAT analysis of variance (ANOVA) program recovers interblock information using ANOVA solutions for the variance components which are the ones usually used in textbooks but are being replaced by some form of maximum likelihood solutions. The code is:

```
"gn-1-0  Triple lattice"
unit [27]
factor  r, b, t
open 'fed933.dat'; channel = 2
read [channel=2]  y, r, b, t;
factor [levels=3] pa, pb, pc
calc pa = newlevels(t; ! ((1...3)3)
&    pb = newlevels(t; ! (3(1...3))
&    pc = newlevels(t; ! (1,2,3, 3,1,2, 2,3,1))
blocks  r/b
treatments  t//(pa+pb+pc)
anova [print=aovt, info, cbmeans; pse=means; fprobability=yes]  y
stop
```

An analysis of variance table with F statistics and probabilities are the same as for the regression analysis. The adjusted means recovering interblock information along with standard errors are obtained, using ANOVA solutions for the variance components.

The GENSTAT variance component procedure uses REML solutions for the variance components to obtain the adjusted treatment means. The variance component solutions are constrained to be positive unless specified otherwise. The program for this analysis is:

```
"gn-1-2.n  Triple lattice"
units [27]
factor r, b, t
open 'fed933.dat'; channel=2
read [channel=2] y, r, b, t;
vcomponents [fixed=t] r/b
reml [print=means, comp,stra; pse=estimates] y
```

```
stop
```

In the vcomponents statement the fixed effect is treatment (T) and the random effects are indicated by R/B. i.e., R and B nested within R are random effects. The output from this program contains the REML solutions for the variance components and their approximate standard errors, a matrix of coefficients of variance components, the adjusted means and their standard error.

The corresponding S-PLUS code for the above analysis is:

```
# SP-1  Triple lattice
#Read data
        dat <- read.table('fed933.dat', col.names=c('y', 'r', 'b', 't'))
        attach (dat)
# Convert r,b,t to factors
        r <- as.factor(r)
        b <- as.factor(b)
        t <- as.factor(t)
# Fit model & print results
        e <- rinfo(y ~ r/b + t, t)
        print ('Variance components & standard errors")
        print (cbind (e$vc,e$se.vc,deparse.level=2))
        print ("REML means & standard errors')
        print (cbind (e$mn,e$se.mn,deparse.level=2))
#Clean up
        detach ()
        rm(dat,r,b,t,e)
```

This program prints out the variance components and their standard errors (questionable) and the REML means and their standard errors. For this particular example, the adjusted means are equal to those obtained from the GENSTAT variance component analysis because the ANOVA and REML solutions for the variance components are equal. In general, they will not be the same but the S-PLUS and SAS PROC MIXED results will be the same.

3. Recovery of Interrow and Intercolumn Information in Lattice Rectangle Designs

The data set used to illustrate the program for recovering interrow and intercolumn information for a lattice rectangle designed experiment is the lattice square design given in Table 12.5 of Cochran and Cox (1957). This data set is given the name lsgr1645.dat in the program. The GENSTAT regression code for a fixed effects analysis is:

```
 "gn-2.1 gn  Lattice square"
units [80]
open 'lsgr1645.dat'; channel = 2
read [channel=2] yield, rep, row, col, grad, treat
model yield
terms rep/(row + col) + treat
fit [print=accumulated; probability=yes] rep/(row + col) + treat
predict treat
```

The GRAD (gradient) column was included in the data file and is used for the analysis in the next section.

The GENSTAT ANOVA program which recovers interrow and intercolumn information is

```
"gn-2-0 gn  Lattice square"
```

```
units [80]
factor yield, rep, row, col, grad, treat
open 'lsgr1645.dat'; channel = 2
read [channel=2] yield, rep, row, col, grad, treat
blocks rep/(row + col)
treatments treat
anova [print=aovt, info, cbmeans; pse=means; fprobability=yes] yield
```

The GENSTAT variance component program for this data set is

```
"gn-2-2 Lattice square"
units [80]
factor rep, row, col, treat
open 'lsgr1645.dat'; channel = 2
read [channel=2] yield, rep, row, col, grad, treat
vcomponents [fixed=treat] rep/(row+col); constraints = positive
reml [print=means,comp,stra; pse=estimates] yield
stop
```

To obtain unconstrained solutions for variance components, simply omit the statement "constraints = positive" in the above.

The corresponding S-PLUS code for recovering interrow and intercolumn information for the above example is

```
#sp-2 Lattice square
#Read data
   dat <- read.table('lsgr1645.dat', col.names=c('yield', 'rep', 'row', 'col', 'gard', 'treat'))
   attach(dat)
#Convert rep,row,col,treat to factors
   rep <- as.factor(rep)
   row <- as.factor(row)
  col <- as.factor(col)
treat <- as.factor(treat)
#Fit model & print results
   e <- rinfo(yield ~ rep/(row+col) + treat, treat)
```

The adjusted means and variance components obtained here agree with SAS PROC MIXED but not with the preceding GENSTAT results.

4. Recovery of Interblock, or Interrow, and Intergradient Information in Incomplete Block or Lattice Rectangle Designs

The data used to illustrate the recovery of interrow and intergradient information are those of the previous section where differential linear gradients within each row are considered rather than considering column effect within each complete block. The values included in the grad data column are the centered linear regression coefficients of position. The gradients and rows are considered to be random effects. Note that this analysis is equally applicable to an incomplete block design with differential gradients within each incomplete block. This form of spatial analysis may be a more appropriate analysis than the standard textbook one. The form of the gradient is linear but additional polynomial or other terms may be added if deemed appropriate.

A GENSTAT regression program for fixed effects is

```
units [80]
factor rep, row, treat
open 'lsgr1645.dat'; channel = 2
read [channel=2] yield, rep, row, col, grad, treat
model yield
terms rep/row/grad + treat
fit [print=accumulated; fprobability=yes] rep/row/grad + treat
predict treat
stop
```

A GENSTAT ANOVA program for this data set is

```
units [80]
factor rep, row, col, treat
open 'lsgr1645.dat'; channel = 2
read [channel=2] yield. rep, row, col, grad, treat
blocks rep/(row+col)
treatments treat
anova [print=aovt,info,cbmeans; pse=means; fprobability=yes] yield
stop
```

A GENSTAT variance component code for REML solutions with positive restraints on the solutions is

```
units [80]
factor rep, row, treat
open 'lsgr1645.dat'; channel = 2
rear [channel = 2] yield, rep, row, col, grad, treat
vcomponents [fixed=treat] rep/row/grad; constraints = positive
read [print=means, comp, stra; pterms=treat; pse=estimates] yield
stop
```

If no restraints are to be placed on variance component solutions, omit the statement "constraints = positive".

The S-PLUS code for recovering interrow and intergradient information for the above example is

```
#sp-3 Lattice square with gradients
#Read data
   dat <- read.table('lsgr1645.dat', col.names=c('yield', 'rep', 'row', 'col', 'grad', 'treat'))
   attach (dat)
#Convert rep,row,treat to factors
   rep <- as.factor(rep)
   row <- as.factor(row)
  treat <- as.factor(treat)
#Fit model & print results
   e <- rinfo(yield ~ rep/row/grad+treat, treat)
  print ('Variance components')
  print (e$vc)
  print ("REML means & standard errors')
  print (cbind(e$mn,e$se.mn,desparse.level=2))
#Clean up
  rm(dat,rep,row,treat,e)
```

Function RINFO is called the model formula YIELD~REP/ROW/GRAD+TREAT where YIELD is the response variate, REP is replicate, ROW is row, GRAD is gradient, TREAT is treatment, and REP/ROW/GRAD indicates that rows are nested within replicates, and gradients are nested within rows.

.

5. Some Comments

GENSTAT regression and SAS PROC GLM programs produce the intra-effect or fixed-effect analyses and give the same results. GENSTAT ANOVA and VCOMPONENTS produce the same adjusted means and variance component solutions but different standard errors. SAS PROC Mixed differs from the preceding two programs in adjusted means, variance component solutions, and standard errors. If no constraints are put on the variance component solutions, then the three programs produce the same results. The adjusted means for S-PLUS agree with SAS but not with GENSTAT output when positive restraints are used. The results from the S-PLUS programs are the same as for SAS. The question of which REML solution for variance components to use is unresolved. Technical Reports (Barnard and Federer, 1996; Federer, 1995) illustrating the programs with examples and annotated computer outputs are available for individuals requiring more than the above programs.

6. Literature Cited

Barnard, J. and W. T. Federer (1996). GENSTAT and SAS programs for recovering interblock, interrow-column, and intergradient information. BU-134?-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, NY 14853. ($5.00)

Cochran, W. G. and G. M. Cox (1957). *Experimental Designs*, second edition. John Wiley & Sons, Inc., New York.

Federer, W. T. (1955). *Experimental Design--Theory and Application.* Macmillan Company, New York (Republished by Oxford & IBH Publishing Company, New Delhi, Bombay, 1967 and 1974)

Federer, W. T. (1995). SAS for ANOVAs and recovery of interblock, intercolumn and intergradient information. BU-1293-M in the Technical Report Series of the Biometrics Unit, Cornell University, Ithaca, New York 14853. ($5.00)

Federer, W. T. and R. D. Wolfinger (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.

GENSTAT 5 Committee (1993). *Genstat 5 Release 3 Reference Manual.* Clarendon Press, Oxford UK.