

RP 02713

STATISTICS UNIT

Report No. 1/84

GENSTAT Sample Runs



ICRISAT

**International Crops Research Institute for the Semi-Arid Tropics
ICRISAT Patancheru P. O.
Andhra Pradesh 502 324, India**

CONTENTS

	Page No.
Introduction	1
Section 1: Logical structure of a Genstat program	2
Section 2: (1) Converting a CRISP file into a text file	5
(2) Creating text files	5
(3) Running a Genstat job	6
(3.1) Running the job on the terminal	6
(3.2) Running a Genstat job in batch system	6
(3.3) Running a Genstat job with a Genstat macro	7
Section 3: Genstat sample programs	
Completely Randomized Design	9
Randomized Block Design	10
Factorial Design	11
Split-Plot Design	12
Strip-Split-Plot Design	13
Strip-Plot or Split-Block Design	14
Latin Square Design	15
Genstat Macros for Lattice Design	16
Square Lattice Design	17
Lattice Square Design	18
Rectangular Lattice Design	19
Covariance Analysis	20
Graphs	21
Histograms	22
Regression Analysis - Simple Linear Regression	23
Multiple Linear Regression	24
References	25

INTRODUCTION

With the sudden change to using Genstat for data analysis we aim in this report to indicate initially how Genstat programs are structured, how to create text files containing programs and data, how to convert CRISP data files to text files, then creating the necessary Genstat programs; how to run these programs on the VAX computer using examples of Genstat programs and macros for the processing of most designed trials. At a later stage we will incorporate more procedures.

Section 1 gives a brief introduction to the logical structure of a Genstat program and we recommend scientists to consult the book by Alvey et al 'Introduction to Genstat', and the Genstat manual, volume 1, for more details.

Section 2 contains:

- (1) the procedures for converting CRISP data files to text files;
- (2) procedures for creating text files (These two procedures are given in more detail in computer services bulletins CS 007, CS 011);
- (3) how to run Genstat programs from the terminal;
- (4) how to run Genstat programs using batch.

Section 3 gives Genstat programs showing examples of frequently-encountered statistical analyses. For each example we give:

- (1) a brief description of the experiment;
- (2) the Genstat program/or macro;
- (3) the computer printouts.

* The computer printouts of the sample runs are available with program offices and statistics unit for consultation.

Section 1

LOGICAL STRUCTURE OF A GENSTAT PROGRAM

All Genstat programs consist of *directives* set between 2 primes. The *first* program *directive* is 'REFERENCE' which should be followed by a name for the program. The *last* program *directive* is 'CLOSE' which as the name implies closes that particular program. To finish using Genstat the *directive* 'STOP' follows 'CLOSE'.

A simple example will illustrate the necessary *directives* to run a Genstat program: yield recorded from a factorial trial with two factors, N with 3 levels, P with 4 levels and 2 blocks (total of 24 plots).

To enable Genstat to know the number of values for each recorded 'VARIATE' and derived 'VARIATES' the *second* program *directive* must declare that number. In this case

```
'UNIT' $ 24
```

Genstat needs to know what 'FACTORS' there are and their respective levels.

```
'FACTOR' N $ 3
'FACTOR' P $ 4
```

implies 'FACTOR' N with 3 levels and P with 4 levels.

With 2 Blocks in the design we also need

```
'FACTOR' BLOCKS $ 2
```

A simple data matrix consists of columns for each 'FACTOR' and 'VARIATE' giving 'FACTOR' levels and 'VARIATE' values under the appropriate column.

To 'READ' the data matrix the 'FACTOR' and 'VARIATE' names are required

```
'READ' BLOCKS, N, P, YIELD
```

The data matrix in this case has 4 columns, the first for BLOCKS, second N, third P, and fourth YIELD.

For Genstat to recognise the End Of Data, a program *directive* 'EOD' follows the data matrix.

For using the analysis of variance section, 3 directives 'BLOCK', 'TREATMENT' and 'ANOVA' are required. It is necessary to define these block and treatment structures for a given design. With the factorial example above we only have blocks, defined by the *directive*

'BLOCK' BLOCKS

The treatments consist of the main effects of N and P and their interaction N.P. Using Genstat terminology $N*P = N + P + N.P$ so that we can define our 'TREATMENT' effects as

'TREATMENT' N*P

We now want the ANOVA with means, standard errors, C.V.; etc. for our variate. This is available with the *directive*

'ANOVA' YIELD

The 'RUN' *directive* enables all the above *directives* to be run.

The complete Genstat program in the example given above is as follows:

```
'REFERENCE' FACTORIAL
'UNITS' $ 24
'FACTOR' N $ 3
'FACTOR' P $ 4
'FACTOR' BLOCKS $ 2
'READ' BLOCKS, N, P, YIELD
'BLOCK' BLOCKS
'TREAT' N*P
'ANOVA' YIELD
'RUN'
```

```
data
```

```
'EOD'
'CLOSE'
'STOP'
```

Section 2

(1) CONVERTING A CRISP FILE INTO A TEXT FILE

\$ CRISP

Enter CRISP option or Utility Option? TEXTFILE

Enter your file name? [CRISP file name XXXXXXXX.DAT]

Enter a name for your Text file? (Name for a text file)

Do you need a list of variables? Yes/No

Enter the range of variables to be stored in the text file?
1-2

Enter the specifications for each variables?

Var 1?

Var 2?

Enter CRISP option or Utility Option? EXIT

\$

(2) CREATING TEXT FILES

We refer to computer services bulletin No. CS 007 and CS 011 for creating text files.

(3) RUNNING A GENSTAT JOB

Once the data and Genstat directive have been stored in a text file, there are several ways of running the job, inspecting the output and saving or printing it.

(3.1) RUNNING THE JOB ON THE TERMINAL

Note that after all the commands given below <CR> means that you must type "return"; the job will not have been completed until the \$ prompt appears on the screen.

Step 1 :- Run the job and check whether there are errors in the input file using

```
$ GENSTAT input file name <CR>
```

and output appears only on the screen.

Step 2 :- If there are no error messages, you must re-run the job and send the output to either the "A" or the "C" line printer using

```
$ GENSTAT/OUTPUT=LPAg:output file name input file name <CR>
```

or

```
$ GENSTAT/OUTPUT=LPCg:output file name input file name <CR>
```

If you would like to save a copy of the output in your account use

```
$ GENSTAT/OUTPUT=output file name input file name <CR>
```

(3.2) RUNNING A GENSTAT JOB IN BATCH SYSTEM

When you run a Genstat job in batch mode, it is not possible to see the output on the screen or to save the output file in your account: output is send straight to the "A" printer only. The command which has to be typed is the following:

```
$ GENSTAT B [input file name] <CR>
```


(3.3) RUNNING A GENSTAT JOB WITH A GENSTAT MACRO

Macros are stored in a permanent file named as **GMACRO**. In order to access the particular macro that you need, you must type the following and answer the prompts that will appear on the screen:

```
$ GENSTAT/OUTPUT=output file name <CR>  
Name of the directive file? input file name <CR>  
Secondary input file? <CR>  
Secondary output file? <CR>  
Userfile 1? GDIR:GMACRO <CR>  
Userfile 2? <CR>
```

GENSTAT SAMPLE PROGRAMS

COMPLETELY RANDOMISED DESIGN

DESCRIPTION: Grain yield of rice using 5 different types of insecticides each replicated 4 times in a completely randomized design.

```
'REFERENCE' COMPLETELY_RANDOMIZED
'UNIT' $ 20
'FACTOR'   TREATMENT $ 5
'READ'     TREATMENT,YIELD
'TREAT'    TREATMENT
'ANOVA'    YIELD
'RUN'
  1  2537
  1  2069
  1  2014
  1  1797
  2  3366
  2  2591
  2  2211
  2  2544
  3  2536
  3  2459
  3  2827
  3  2385
  4  2387
  4  2453
  4  1556
  4  2116
  5  1997
  5  1679
  5  1649
  5  1859
'EOD'
'CLOSE'
'STOP'
```

RANDOMISED BLOCK DESIGN

DESCRIPTION: Grain yield of rice variety planted at 6 different seeding rates in 4 blocks. The factors involved are seed rate and blocks.

```
'REFERENCE' RANDOMISED_BLOCK_DESIGN
'UNIT' # 24
'FACTOR'     BLOCKS # 4
'FACTOR'     SEEDRATE # 6
'READ'       BLOCKS,SEEDRATE,YIELD
'BLOCK'      BLOCKS
'TREATMENT'  SEEDRATE
'ANOVA'      YIELD
'RUN'
  1  1  5113
  1  1  5324
  1  1  5272
  1  1  5164
  1  1  4804
  1  1  5254
  2  1  5398
  2  1  5952
  2  1  5713
  2  1  4831
  2  1  4848
  2  1  4542
  3  1  5307
  3  1  4719
  3  1  5483
  3  1  4986
  3  1  4432
  3  1  4919
  4  1  4678
  4  1  4264
  4  1  4749
  4  1  4410
  4  1  4748
  4  1  4098
'EOD'
'CLOSE'
'STOP'
```

FACTORIAL DESIGN ANALYSIS

DESCRIPTION: Yield from a factorial experiment in 3 blocks with 2 factors, nitrogen and phosphate each at 2 levels.

Note: The same program can be used for more factors by changing the 'TREATMENT' model and including the additional factor names and levels. Example for a 3 factor factorial with N,P,K as factors - 'TREATMENT' N*P*K.

```
'REFERENCE' FACTORIAL_EXPERIMENT
'UNIT' $ 12
'FACTOR'     REP $ 3
'FACTOR'     N $ 2
'FACTOR'     P $ 2
'READ'       REP,N,P,YIELD
'BLOCK'      REP
'TREATMENT'  N*P
'ANOVA'      YIELD
'RUN'
  1          1    39.50
  2          1    43.10
  3          1    45.20
  1          1    38.60
  2          1    39.50
  3          1    33.00
  1          2    27.20
  2          2    23.20
  3          2    24.80
  1          2    24.60
  2          2    24.20
  3          2    22.20
'EOD'
'CLOSE
'STOP'
```

SPLIT PLOT DESIGN

DESCRIPTION: Yield of 3 varieties of Alfalfa with 4 dates of harvesting in 2 blocks from a split plot trial with varieties as main plot treatments and dates of harvest as sub plot treatments.

Note: The same directives can be used for a split split plot analysis where 3 treatment factors are involved by changing the 'BLOCK' and 'TREATMENT' models. See the example of a split split plot analysis which follows.

```
'REFERENCE' SPLIT_PLOT
'UNIT' $ 24
'FACTOR'    BLOCK $ 2
'FACTOR'    DATES $ 4
'FACTOR'    VARIETY $ 3
'READ'      BLOCK,DATES,VARIETY,YIELD
'BLOCK'     BLOCK/VARIETY/DATES
'TREATMENT' VARIETY*DATES
'ANOVA'     YIELD
'RUN'
  1         1         2.17
  2         1         1.88
  1         1         1.58
  2         1         1.26
  1         1         2.29
  2         1         1.60
  1         1         2.23
  2         1         2.01
  1         2         2.33
  2         2         .01
  1         2         .38
  2         2         .30
  1         2         .86
  2         2         .70
  1         2         .27
  2         2         .81
  1         3         .75
  2         3         .95
  1         3         .52
  2         3         .47
  1         3         .55
  2         3         .61
  1         3         .56
  2         3         1.72
'EOD'
'CLOSE'
'STOP'
```

SPLIT SPLIT PLOT DESIGN

DESCRIPTION: Grain yield data from a split split plot experiment with 3 varieties of rice (sub sub plot treatment) grown under 3 management practices (sub plot treatment) at 5 levels of nitrogen fertiliser (main plot treatment) in 3 blocks.

```
'REFERENCE' SPLIT_SPLIT_PLOT
'UNIT'      $ 135
'FACTOR'    BLOCK $ 3
'FACTOR'    NITROGEN $ 5
'FACTOR'    MANAGEMENT $ 3
'FACTOR'    VARIETY $ 3
'READ'      BLOCK,NITROGEN,MANAGEMENT,VARIETY,YIELD
'BLOCK'     BLOCK/NITROGEN/MANAGEMENT/VARIETY
'TREATMENT' NITROGEN*MANAGEMENT*VARIETY -
'ANOVA'     YIELD
'RUN'
  1      1      1      1      3.320
  1      1      2      1      3.766
  1      1      3      1      4.660
  1      2      1      1      3.188
  1      2      2      1      3.625
      ..
      ..
  3      4      2      3      7.778
  3      4      3      3      9.128
  3      5      1      3      8.032
  3      5      2      3      9.294
  3      5      3      3      9.712
'END'
'CLOSE'
'STOP'
```

STRIP PLOT DESIGN

DESCRIPTION: Data on grain yields of 3 rice varieties grown under 3 nitrogen levels from a strip plot trial in 3 blocks. The factors are variety and nitrogen.

```
'REFERENCE' STRIP_PLOT_ANALYSIS
'UNIT' $ 27
'FACTOR'     BLOCK $ 3
'FACTOR'     NITROGEN $ 3
'FACTOR'     VARIETY $ 3
'READ'       BLOCK,NITROGEN,VARIETY,YIELD
'BLOCK'      BLOCK/(NITROGEN*VARIETY)
'TREATMENT'  NITROGEN*VARIETY
'ANOVA'      YIELD
'RUN'
  1  1  1  2373
  1  2  1  4076
  1  3  1  7254
  2  1  1  3958
  2  2  1  6431
  2  3  1  6808
  3  1  1  4384
  3  2  1  4889
  3  3  1  8582
  1  1  2  4007
  1  2  2  5630
  1  3  2  7053
  2  1  2  5795
  2  2  2  7334
  2  3  2  8284
  3  1  2  5001
  3  2  2  7177
  3  3  2  6297
  1  1  3  2620
  1  2  3  4676
  1  3  3  7666
  2  1  3  4508
  2  2  3  6672
  2  3  3  7328
  3  1  3  5621
  3  2  3  7019
  3  3  3  8611
'EOD'
'CLOSE'
'STOP'
```


LATIN SQUARE DESIGN

DESCRIPTION: Grain yield of 3 promising corn hybrids and a check variety from a latin square design in 4 rows and 4 columns. The blocking factors are rows, columns and treatment factor is variety.

```
'REFERENCE' LATIN_SQUARE
'UNIT' $ 16
'FACTOR' ROWS $ 4
'FACTOR' COLUMN $ 4
'FACTOR' VARIETY $ 4
'READ' ROWS,COLUMN,VARIETY,YIELD
'BLOCK' ROWS*COLUMN
'TREATMENT' VARIETY
'ANOVA' YIELD
'RUN'
  1  1  2  32.8
  1  2  4  24.2
  1  3  3  28.5
  1  4  1  26.9
  2  1  3  29.5
  2  2  1  23.7
  2  3  4  28.0
  2  4  2  25.8
  3  1  1  33.4
  3  2  3  14.2
  3  3  2  33.3
  3  4  4  23.6
  4  1  4  31.3
  4  2  2  25.8
  4  3  1  33.1
  4  4  3  13.2
'EOD'
'CLOSE'
'STOP'
```

G E N S T A T M A C R O S
F O R L A T T I C E D E S I G N S

It is easy to obtain the 'intra-block' analysis described in Cochran and Cox chapter 9, using the 'BLOCK', 'TREAT' and 'ANOVA' directives of Genstat. However, the 'intra-block' analysis ignores the information about the treatments which is contained in the block totals (Yates, 1940). The following macros give the combined analysis for all possible lattice designs, incorporating the recovery of inter-block information by means of pseudofactors (See Genstat Manual Ch.6:8)

**SQUARE LATTICE DESIGNS
(SQLAT)**

This macro analyses square lattice designs with any number of replications and repeats of the design (Cochran and Cox, 1957, ch.10).

Method of use:-

The user must provide 5 factors which specify the design and a set of variates to be analysed. The identifiers are:

SREP	factor	Repeat of the design
UREP	factor	Replications
BWR	factor	Blocks within reps
PLOT	factor	Plots
TREAT	factor	Treatment factor
YSET	set	A set of variates to be analysed

Factor PLOT shows the position of the treatment applied in a block within a replication.

NOTE:- In the 'REFERENCE' statement the options MUNN (number of unnamed structures) and NID (number of identifiers) may be necessary to increase the work space.

Yield data from a 7 x 7 triple lattice millet trial conducted in Senegal.

```
'REFERENCE' SQUARE_LATTICE
'UNIT' $ 147
'FACTOR' SREP $ 1
'FACTOR' UREP $ 3
'FACTOR' BWR $ 7
'FACTOR' PLOT $ 7
'FACTOR' TREAT $ 49
'READ' SREP,UREP,BWR,PLOT,TREAT,YIELD
'GET/FILE=1' SQLAT$SQLAT
'SET' YSET=YIELD
'USE/R' SQLAT$
'RUN'
  1  1  2  1  14  4006
  1  1  2  2  10  1700
  1  1  2  3  13  3211
  1  1  2  4  9  2014
      ..
      ..
  1  3  6  5  31  4002
  1  3  6  6  6  2769
  1  3  6  7  39  1690
'EOD'
'CLOSE'
'crop'
```

The analysis of Lattice Square Designs
(LATTSQ)

This macro analyses lattice square designs with any number of replicates and repeats of the basic design.

Method of use:-

The user must provide 5 factors which specify the design and a set of variates to be analysed. The identifiers are

SREP	factor	Repeats of the design
UREP	factor	Replicates
RWR	factor	Rows within reps
CWR	factor	Columns within reps
TREAT	factor	Treatments
YSET	set	A set of variates to be analysed

Yield data from a 5 x 5 balanced lattice square design in 3 replications (Cochran and Cox p.486).

NOTE:- In the 'REFERENCE' statement, options NUNN=300 and NID=300 are included to increase the workspace available.

```
'REFERENCE/NUNN=300,NID=300 LATTICE_SQUARE
'UNIT' $ 75
'FACTOR' SREP $ 1
'FACTOR' UREP $ 3
'FACTOR' RWR $ 5
'FACTOR' CWR $ 5
'FACTOR' TREAT $ 25
'READ SREP,UREP,RWR,CWR,TREAT,YIELD
'GET/FILE=1' LATTSQ$LATTSQ
'SET' YSET=YIELD
'USE/R' LATTSQ$
'RUN'
  1  1  1  1  18  33.3
  1  1  1  2  9  30.
  1  1  1  3  11  35.
  1  1  1  4  2  30.1

  1  3  5  2  4  32.8
  1  3  5  3  12  31.9
  1  3  5  4  25  28.8
  1  3  5  5  16  30.6
'EOD'
'CLOSE'
'CTOP'
```

**The analysis of lattice designs
(LATTICE)**

The macro analyses either square or rectangular lattice designs with any number of replicates and repeats of the basic design.

Methods of use:-

The user must provide 4 factors which specify the design and a set of variates to be analysed. The global identifiers are:-

SREP	factor	Repeats of the design
UREP	factor	Replications
BWR	factor	Blocks within reps
TREAT	factor	Treatment factor
YSET	set	A set of variates to be analysed

NOTE:- In the 'REFERENCE' statement the options NUNN (number of unnamed structures) and NID (number of identifiers) may be necessary to increase the work space.

N.B. For rectangular lattice designs, BWR must be ordered by block partners (Cochran and Cox, 1957, p. 417)

```
'REFERENCE'  RECTANGULAR_LATTICE
'UNITS' $ 36
'FACTOR'    SREP $ 1
'FACTOR'    UREP $ 3
'FACTOR'    BWR $ 4
'FACTOR'    PLOT $ 3
'FACTOR'    TREAT $ 12
'READ'      UREP,SREP,BWR,PLOT,TREAT,YIELD
'SET'       YSET=YIELD
'GET/FILE=1' LATTICE$LATTICE
'USE/R'     LATTICE $
'RUN'
  1      1      1      1      2      9
  1      1      1      2      3      4
  1      1      1      3      1     16

  3          3      3      4      1
  3          4      1      5     11
  3          4      2      1     22
  3          4      3      7     17

'EOD'
'CLOSE'
'STOP'
```

COVARIANCE ANALYSIS

DESCRIPTION: Millet yield data from a split split plot design with 4 varieties (main plot treatment), 3 levels of fertiliser (sub plot treatment) and 2 levels of plant density (sub sub plot treatment) in 6 blocks. The covariate is the distance of each plot from a line of trees at one side of the field.

Covariate analysis will be carried out providing that the directive 'COVARIATE', followed by the list of covariates, precedes the 'ANOVA' statement. Covariate analysis can be done for any type of design and for any number of covariates.

```
'REFERENCE' COVARIANCE_ANALYSIS
'UNIT' $ 144
'FACTOR' REP $ 6
'FACTOR' PERT $ 3
'FACTOR' DENSITY $ 2
'FACTOR' ENTRY $ 4
'READ' REP,PERT,DENSITY,ENTRY,YIELD,DISTANCE
'BLOCK' REP/ENTRY/PERT/DENSITY
'TREATMENT' ENTRY*PERT*DENSITY
'COVARIATE' DISTANCE
'ANOVA' YIELD
'RUN'
  1  1  2  3  2176  1
  1  1  1  3  2030  1
  1  3  1  3  1726  1
  1  3  2  3  2186  1
      ..
      ..
  6  2  1  1  4535  6
  6  2  2  1  4654  6
  6  3  2  1  4597  6
  6  3  1  1  5136  6
'EOD'
'CLOSE'
'STOP'
```

PLOTTING DATA IN GRAPHS

Plotting of grain yield of rice variety Taichung Native 1 against number of tillers.

```
'REFERENCE' GRAPH
'UNIT' $ 8
'READ' YIELD,TILLER
'GRAPH' YIELD,TILLER
'RUN'
  5380 293
  5510 325
  6000 332
  5840 342
  6416 342
  6666 378
  7016 380
  6994 410
'EOD'
'CLOSE'
'STOP'
```

H I S T O G R A M S

Histogram showing the distribution of plant height from 25 hybrid plants.

```
'REFERENCE' HISTOGRAM
'UNIT' $ 25
'READ' HEIGHT
'HISTOGRAM' HEIGHT
'RUN'
  0.55
  1.10
  1.09
  1.22
  1.26
  1.06
  1.24
  0.25
  0.53
  1.10
  1.35
  0.15
  0.26
  0.43
  1.14
  1.15
  1.23
  1.18
  1.29
  1.17
  1.15
  0.81
  1.10
  1.45
  0.31
'EOD'
'CLOSE'
'STOP'
```


SIMPLE LINEAR REGRESSION ANALYSIS

DESCRIPTION: Computation of a linear regression equation between grain yield and number of tillers for rice variety Taichung Native 1. The 'TERMS' directive specifies both dependent and independent variables. The 'Y' directive indicates the dependent variable and the 'FIT' directive the independent variable.

```
'REFERENCE' LINEAR_REGRESSION
'UNIT' $ 8
'READ' YIELD,TILLERS
'TERMS' YIELD,TILLERS
'Y' YIELD
'FIT' TILLERS
'RUN'
  5380 293
  5510 325
  6000 332
  5840 342
  6416 342
  6666 378
  7016 380
  6994 410
'EOD'
'CLOSE'
'STOP'
```

MULTIPLE LINEAR REGRESSION ANALYSIS

DESCRIPTION: Grain yield data of 8 rice varieties is to be regressed on plant height and the number of tillers/hill. The 'TERMS' directive must include all 3 variables. The 'Y' directive defines grain yield as the dependent variable and the 'FIT' directive defines two independent variables, plant height and tillers.

```
'REFERENCE' MULTIPLE_REGRESSION
'UNIT' $ 8
'READ' YIELD,HEIGHT,TILLERS
'TERMS' YIELD,HEIGHT,TILLERS
'Y' YIELD
'FIT' HEIGHT,TILLERS
'RUN'
  5755  110.5  14.5
  5939  105.4  16.0
  6010  118.1  14.6
  6545  104.5  18.2
  6730   93.6  15.4
  6750   84.1  17.6
  6899   77.8  17.9
  7862   75.6  19.4
'EOD'
'CLOSE'
'STOP'
```

REFERENCES

- Alvey,N., Galwey,N., and Lane,P. 1982. An introduction to Genstat. Academic Press, London, New York.
- Cochran,W.G., and Cox,M.G. 1957. Experimental Designs. John Wiley and Sons, New York.
- Genstat (Release 4.04) Part I. 1983. Rothamsted Experimental Station, UK.
- Venkateswarlu,P., and Gilliver,B. 1982. Statistics Unit Report No. 9/82.
- Gomez,K.A., and Gomez,A.A. 1981. Statistical procedures for agricultural research. The International Rice Research Institute, Manila, Philippines.
- Yates,F. 1940. The recovery of inter-block information in balanced incomplete block designs. Annals of Eugenics, 10, 317-325.