Use of Statistical Package SAS
A Case Study from Coconut Research Institute of Sri Lanka

T. S. G. PEIRIS
Coconut Research Institute, Lunuwila, Sri Lanka

and

R. D. STERN
Department of Applied Statistics, University of Reading, Reading, England

ABSTRACT

The Coconut Research Institute in Sri Lanka (CRI) began using microcomputers for the analysis of their experimental data in 1986. Four case studies of typical sets of field experimental data from the CRI are used to illustrate the potential benefits and possible problems in moving from a manual system of data processing to one which uses a powerful statistical package.

INTRODUCTION

The Coconut Research Institute of Sri Lanka (CRI) acquired microcomputers in 1986 for data processing and analysis. Other agricultural research institutes in Sri Lanka are also currently acquiring microcomputers and all have a joint licence to use the statistical package SYSTEM ANALYSIS SOFTWARE (SAS). This paper considers some ways in which computerization may affect the biometrical work at these institutes.

The structure of the CRI, collection of data from field experiments and their analyses are briefly described. Then four typical sets of experimental data and their analyses are outlined. The emphases in this paper is on the potential of the computing facilities to enhance the work done by the Biometry Unit, rather than the sets of data themselves. Some of the dangers of uncritical use of a package for the analysis are also considered.

In the final section of the paper we consider briefly other ways in which the computer could help in data processing. Many agricultural research institutes conduct experiments for more than one year. This is however taken to extremes with a crop like coconut where palm yields are taken six times a year for trees which yield for over 60 years. The entry and management of such long term data are a formidable task which could involve other software in addition to a statistical package.

BACKGROUND

The Coconut Research Institute (CRI) at Lunuwila in the North Western Province in Sri Lanka was established in 1929. It is primarily concerned with research on agronomic aspects of coconut, on the processing of coconut and its byproducts and conducts a large number of field and laboratory experiments. The field experiments are carried out either at the main station (Bandirippuwa Estate) or at one of the eight substations or in state-owned estates. Most of the field experiments are essentially long-term, and on an average, eight new field experiments are initiated annually.
The CRI has eight research divisions but most of the field experiments are conducted by Agronomy, Soils and Plant Nutrition, Genetics and Plant Breeding and Crop Protection Divisions. The research staff number about 40 and all divisions have officers with postgraduate degrees.

The Biometry Unit at the CRI is the oldest biometry facility in the research institutes in Sri Lanka, and its responsibility is to provide a statistical service to the other research divisions. This involves the design of experiments, analysis of data and the interpretation of results in consultation with the experimenter concerned. In addition, the Biometry Unit conducts research on crop-weather relationships and maintains three agrometeorological stations. The Unit has two biometricians with a number of support staff.

Normally, the Biometry Unit designs experiments in consultation with the experimenter concerned. The most commonly used design is the randomised block design with factorial treatment structure. The number of blocks is usually two or three with plots often consisting of six to eight coconut palms.

In field experiments, coconuts are harvested once every two months, and at each harvest two bunches are picked. The common variables recorded are the number of nuts and their copra content (kernel dried to 6% moisture) and the number of female flowers. The data are usually recorded for individual palms. After each pick the data are summarised over plots and at the end of the last pick, the data are totalled over the year. The original data, recorded in the field books, are transferred to record books. The yearly summary data are kept in files.

These summary data are then analysed by the staff in the Biometry Unit. The methods of analysis are standard, involving the analysis of variance, covariance and regression analysis. The data are sometimes transformed before analysis. Most analyses, regardless of the size of the data, are carried out with big electric desk calculators. Complicated analyses such as principal components analysis and multiple regression studies are sent to the computer at the Statistics and Computer Science Department in the University of Colombo.

The summary of the analysis is returned to the experimenter with the data. This summary usually consists of the analysis of variance table together with the means and standard errors etc. The method of the analysis and the interpretation of the results are explained verbally to the experimenter. Occasionally a brief written report is prepared.

In mid-1986, two microcomputers were obtained for the Biometry Unit. Until the SAS package was acquired, the support staff who were not familiar with computers, were trained in the general use of computers. They also entered and analysed a few sets of data with a simple statistical package provided by manufacturer.

MATERIALS AND METHODS

Data sets used in the Study

Four sets of experimental data have been selected from three research divisions to compare the analysis with and without the use of the computer. They are typical of the experiments conducted by the CRI. A brief summary of the experiments is given in Fig. 1.
Fig. 1. Brief summary of the experiments

| Example | Object | Treatment | Expt. design | Number of blocks | Plot size | Character analysed |
|---------|--------|-----------|--------------|------------------|-----------|--------------------|
| 1       | Study of fertilizer effect on coconut | nitrogen and potash each at 3 levels. | RBD | 3 | 8 palms | yield per plot |
| 2       | Evaluation of 3 legume species each with 3 varieties grown in 5 different shade levels | main: 5 shade levels, sub: variety legume. | split plot design. | 3 | 5 plants | seed yield per plot |
| 3       | Comparison of 3 nitrogen sources. | 2 levels of each source and control. | RBD | 3 | 12 palms | yield per plot |
|         | Study of systemic insecticides for the control of black beetle in coconut seedlings | seven insecticides and control. | RBD | 8 | 1 palms | % of fronds attacked |

The first experiment had been conducted for eight years by the Soils and Plant Nutrition Division in an estate in Chilaw district. It is a fertilizer experiment with two quantitative factors consisting of all combinations of three levels of nitrogen and potash. The layout is a randomised block design with eight palms in each plot and three blocks. The nut yield and copra content had been collected. The yield in the first year is used for illustration.

The second experiment had been conducted in a shade house by the Agronomy Division to evaluate three varieties of each of three legumes: cowpea, black gram and green gram, each grown under five different shade levels. The layout is a split-plot design in three randomised blocks. The five shade levels are the main plots and nine combinations of legume varieties are the sub plots. A large number of variables were measured. The seed yield is analysed for illustration.
A Case Study from Coconut Research Institute of Sri Lanka

The third experiment had been conducted for nine years by the Soils and Plant Nutrition Division in an estate in Chilaw district and is a comparison of three sources of nitrogen: urea, ammonium chloride and ammonium sulphate, each of which had been applied at two levels. There was also a control treatment with no added nitrogen; replicated three times within each block. The treatment structure is therefore a simple factorial with unequal replication for the control. The analysis described in this example is an analysis of covariance with the yield of the final year as the variable and the yield in the one year before the treatments as a covariate. The analysis of covariance is often used in the analysis of coconut data.

These three examples are typical of much of the current work of the Biometry Unit.

The final example involves data for individual palms. The volume of data is large. The problems of manual analysis are therefore correspondingly greater and there is more potential for improvement. This, therefore, typifies the type of data where the computer should have a greater impact. This experiment had been carried out by Crop Protection Division at Bandirippuwa estate to evaluate insecticides for the control of rhinoceros beetle which attacks the fronds of coconut palms. Seven insecticides and a control were used in a randomized block design with block of eight palms. The insecticides were applied once. The total number of coconut fronds and those attacked were counted at two-month intervals. The first measurement had been taken before the insecticides were applied.

RESULTS

Analysis of the four data sets

In this section the manual analysis is compared with the results from the use of SAS. Two SAS procedures are used. PROC ANOVA handles balanced designs while PROC GLM (General Linear Model) can be used whether the design is balanced or not. When it is appropriate, PROC ANOVA is faster than PROC GLM and presents the results in a way that can be easier to interpret. However, even for balanced design, some options (such as the division of treatment effects into individual contrasts and covariance analysis) are not available in PROC ANOVA. In such situations PROC GLM has to be used.

Analysis 1. (RBD with factorial structure)

The experimenter was interested in the overall analysis of variance and also in the linear and quadratic components of the main effect plus possible interactions between the two factors. The results as prepared manually by the Biometry Unit are given in Fig. 2. This was a standard routine for the staff and involved no difficulty.

The input of the data and the factor levels to SAS was straight-forward and is shown together with the results from using PROC ANOVA in Fig. 3. Minor disappointment in the presentation of the results are the excess number of figures given for each of the terms and the fact that the two-way table of means is not presented as such.
### Manual analysis for the data set

| Source       | df. | SS    | MS    | F-value |
|--------------|-----|-------|-------|---------|
| Block        | 2   | 5,178 | 2,539 |         |
| Treatments   | 8   | 34,005| 4,251 | 1.86    |
| NI           | 1   | 968   | 968   | 0.42    |
| Nq           | 1   | 10,361| 10,361| 4.53*   |
| NI           | 1   | 11,807| 11,807| 5.16*   |
| Nq           | 1   | 1,483 | 1,483 | 0.65*   |
| NK1l         | 1   | 4,721 | 4,721 | 2.06    |
| NK1q         | 1   | 121   | 121   | 0.05    |
| NK1q1        | 1   | 7     | 7     | 0.0     |
| NK1qq        | 1   | 4,537 | 4,537 | 1.98    |
| Error        | 16  | 36,607| 2,288 |         |
| Total        | 26  | 75,790|       |         |

CV = 8.82%

**Mean values.**

| K level | 0   | 1   | 2   |
|---------|-----|-----|-----|
| 0       | 678.3 | 632.3 | 667.7 |
| 1       | 659.43 | 659.43 | 659.43 |
| 2       | 625.23 | 674.10 | 639.0  |

S.E. for difference between any two means (16 df) = 39.06

N levels:

| 0   | 1   | 2   |
|-----|-----|-----|
| 659.43 | 659.43 | 659.43 |
| 625.23 | 674.10 | 639.0  |

S.E. for difference between any two means (16 df.) = 22.55.

### SAS program and the output for the data set

```sas
DATA KOB 1.SAS;
DO NLEVEL = 0 TO 2;
DO KLEVEL = 0 TO 2;
DO BLOCK = 1 TO 3;
INPUT YIELD;
OUTPUT;
END;
END;
END;
CARDS;
726 621 688 688 592 617 648 659 696
650 709 573 654 675 593 333 637 583
782 712 693 644 636 683 641 700 576
RUN;
PROC ANOVA;
CLASS BLOCK NLEVEL KLEVEL;
MODEL YIELD=BLOCK NLEVEL KLEVEL NLEVEL* KLEVEL;
MEANS NLEVEL KLEVEL NLEVEL* KLEVEL/LSD;
RUN;
```

General Linear Models Procedure

42
A Case Study from Coconut Research Institute of Sri Lanka

Dependent Variable: YIELD

| Source          | DF | Sum of Squares | Mean Square | F Value | Pr—F  |
|-----------------|----|----------------|-------------|---------|-------|
| Model           | 10 | 39,182.370370  | 3,918.237037| 1.71    | 0.1627|
| Error           | 16 | 36,607.481841  | 2,287.967593|         |       |
| Corrected Total| 26 | 75,789.852852  |             |         |       |

R-Square        | 0.516987
CV              | 7.3259006
Root MSE        | 47.832704
Y Mean          | 652.92592593

Dependent Variable: YIELD

| Source       | DF | Type I SS  | Mean Square | F Value | Pr—F  |
|--------------|----|------------|-------------|---------|-------|
| BLOCK        | 2  | 5,177.852  | 2,588.926   | 1.13    | 0.3470|
| NLEVEL       | 2  | 11,329.185 | 5,664.593   | 2.48    | 0.1157|
| KLEVEL       | 2  | 13,289.852 | 6,644.926   | 2.90    | 0.0836|
| NLEVEL*KLEVEL| 4  | 9,385.481  | 2,346.370   | 1.03    | 0.4240|

T tests (LSD) for variable: YIELD

**NOTE:** This test controls the type I comparison-wise error rate not the experiment-wise error rate.

Alpha = 0.05 df=16 MSE = 2287.968
Critical Value of T=2.12
Least Significant Difference = 47.801

Means with the same letter are not significantly different.

| T Grouping | Mean  | N KLEVEL |
|------------|-------|----------|
| A          | 674.11| 9 2      |
| A          | 659.44| 9 0      |
| B A        | 625.22| 9 1      |

**NOTE:** This test controls the type I comparison-wise error rate not the experiment-wise error rate.

Alpha = 0.05 df=16 MSE = 2287.968
Critical Value of T=2.12
Least Significant Difference = 47.801

Means with the same letter are not significantly different.

| T Grouping | Mean  | N XLEVEL |
|------------|-------|----------|
| A          | 683.78| 9 0      |
| A          | 642.44| 9 1      |
| B A        | 632.56| 9 2      |

| Level of NLEVEL | Level of KLEVEL | N | YIELD Mean | SD |
|-----------------|-----------------|---|------------|----|
| 0               | 0               | 3 | 678.333333 | 53.1632705 |
| 0               | 1               | 3 | 632.333333 | 49.8029450 |
| 0               | 2               | 3 | 667.666667 | 25.1462390 |
| 1               | 0               | 3 | 644.000000 | 68.1982404 |
| 1               | 1               | 3 | 640.666667 | 42.5949919 |
| 1               | 2               | 3 | 591.000000 | 42.5675933 |
| 2               | 0               | 3 | 729.000000 | 46.8721666 |
| 2               | 1               | 3 | 654.333333 | 25.1462390 |
| 2               | 2               | 3 | 639.000000 | 62.0241888 |
A more disturbing omission is the lack of the standard errors for the treatment means. The closest option (given in Fig. 3) is to ask for least significant differences (LSD). However, this gives the LSD for the main effects only. This is easy to misinterpret if there is a substantial interaction. In addition to the LSD there are 17 different multiple comparison options. Such tests could easily mislead the experimenter. They still give no measure of precision for the interaction means. The standard deviations given for the interaction means in Fig. 3 should not be used to interpret the results. They are based only on the three observations from treatment combination, while the grand analysis assumes the same variance for all treatments. Since PROC ANOVA does not present the subdivision of the treatment sums of squares PROC GLM were used for this. The additional output is shown in Fig. 4.

Fig. 4. Additional output for data set 1 from PROC GLM.

| Source          | DF | Type III SS | Mean Square | F Value | Pr > F |
|-----------------|----|-------------|-------------|---------|--------|
| Block           | 2  | 5,177.852   | 2,588.926   | 1.13    | 0.3470 |
| NLEVEL          | 2  | 11,329.185  | 5,664.593   | 2.48    | 0.1157 |
| KLEVEL          | 2  | 13,289.852  | 6,644.926   | 2.90    | 0.0839 |
| NLEVEL*KLEVEL   | 4  | 9,385.481   | 2,346.370   | 1.03    | 0.4240 |

Dependent Variable: YIELD

| Contrast        | DF | Contrast SS | Mean Square | F Value | Pr > F |
|-----------------|----|-------------|-------------|---------|--------|
| NL              |    | 968.000000  | 968.000000  | 0.42    | 0.5246 |
| NQ              | 1  | 10,361.185185 | 10,361.185185 | 4.53    | 0.0492 |
| KL              | 1  | 11,806.722222 | 11,806.722222 | 5.16    | 0.0373 |
| KQ              | 1  | 1,483.129630  | 1,483.129630 | 0.65    | 0.4325 |
| NKLL            | 1  | 4,720.333333  | 4,720.333333 | 2.06    | 0.1702 |
| NKLQ            | 1  | 7.1111111111  | 7.1111111111 | 0.00    | 0.9562 |
| NKQL            | 1  | 121.000000   | 121.000000  | 0.05    | 0.8210 |
| NKQQ            | 1  | 4,537.037037  | 4,537.037037 | 1.98    | 0.1782 |

The outputs from these two figures are now quite extensive compared with the information given from the manual analysis in Fig. 2. It is acceptable as a basis but requires cautionary notes and editing before it could be returned to the experimenter. A further point is that the default option for the GLM procedure produces two sets of sums of squares (Type I and III) even for a simple balanced design, where they are identical. To prevent this, the user must know of the theory of General Linear Models.
**A Case Study from Coconut Research Institute of Sri Lanka**

**Analysis 2. (Split plot design)**

Fig. 5 shows the analysis of variance table for this split plot design.

**Fig. 5 Manual analysis for the data set 2**

| Source            | DF | SS     | MS   | F-Value |
|-------------------|----|--------|------|---------|
| Block             | 2  | 0.354  | 0.177|         |
| Sh. level         | 4  | 450.882| 112.721| 1,034.14*|
| Error 1           | 8  | 0.869  | 0.109| 1.31    |
| Varieties         | 8  | 94,024 | 11.753| 141.60*|
| bt. species       | 2  | 84.983 | 42.490| 511.93*|
| var. wt. spec.    | 6  | 9.040  | 1.510| 18.19*  |
| Sh x variety      | 32 | 35.097 | 1.097|         |
| Error 2           | 80 | 6.649  | 0.083|         |
| Total             | 134| 587.875|      |         |

C.V. = 7.1%

**Means**

| Shlevel | 1 | 2 | 3 | 4 | 5 | 6.88 | 5.44 | 3.47 | 2.48 | 2.05 |
|---------|---|---|---|---|---|------|------|------|------|------|

S.E. for difference between means (80d.f.) = 0.090

| Species | 1 | 2 | 3 | 3.34 | 5.18 | 3.68 |
|---------|---|---|---|------|------|------|

S. E. difference between means (80 d.f.) = .061

| Variety | 1 | 2 | 3 |
|---------|---|---|---|
| Species | 1 | 2 | 3 |
| 1       | 3.26 | 3.12 | 3.63 |
| 2       | 5.54 | 4.87 | 5.13 |
| 3       | 4.06 | 3.37 | 3.60 |

S.E. for difference between means (80 d.f.) = .105

The results for PROC ANOVA in SAS together with the SAS program are given in Fig. 6. The MEAN option was omitted in the program because the output was illustrated in the first example. Subdivision of the variety term is not possible in the ANOVA procedure. A more serious problem however is that the presentation of the results for a split-plot experiment could be confusing. It is not satisfactory to return results to experimenter where the main plot term appears twice with identical mean squares but different F values. This is a problem for all analysis that involve different error levels.
T. S. G. PEIRIS and R. D. STERN

Fig. 6 SAS program and the output for the data set 2

DATA MAHL. DATA;
DO BLOCK = 1 TO 3;
DO SHLEVEL = 1 TO 5; VAR = 0;
DO SPECIES = 1 TO 3;
DO SPVAR = 1 TO 3; VAR = VAR + 1;
INPUT SEEDY —; OUTPUT;
END; END; END; END;
CARDS;
5.46  5.40  5.44  8.74  8.86  8.70  6.80  6.33  6.78
4.04  3.90  4.48  6.85  6.30  6.57  6.15  5.11  5.80
3.00  2.85  3.36  5.28  4.06  4.35  3.36  2.30  2.57
2.10  2.05  2.70  3.36  2.89  3.10  2.30  1.61  1.79
1.90  1.71  2.30  3.25  2.20  2.56  2.00  1.43  1.56
5.25  5.52  5.38  8.46  8.61  8.50  6.68  6.40  6.70
4.22  4.05  4.50  6.50  6.02  6.20  6.00  4.93  5.18
3.05  2.96  3.65  4.80  3.62  3.95  3.22  2.22  2.45
2.30  2.25  3.08  3.45  2.30  3.00  2.58  1.72  1.97
2.01  1.94  2.61  2.65  2.02  2.25  2.15  1.36  1.50
4.88  4.97  4.90  9.20  9.10  9.08  6.72  6.60  6.22
4.00  3.85  4.36  7.70  6.75  7.02  5.90  5.18  5.50
2.80  2.77  3.27  6.15  4.55  5.58  3.15  2.40  2.68
2.08  1.85  2.37  3.70  3.40  3.33  2.22  1.68  1.89
1.80  1.48  2.06  2.93  2.36  2.70  1.90  1.29  1.43
RUN;
PROC ANOVA;
CLASS BLOCK SHLEVEL VAR SPECIES SPVAR;
MODEL SEEDY = BLOCK SHLEVEL BLOCK*SHLEVEL VAR SHLVL*VAR;
TEST H= BLOCK SHLEVEL E= BLOCK*SHLEVEL;
RUN;

Defendent Variable: SEEDYY

| Source          | DN | Some of Squares | Mean Square | F Value | Pr > F |
|-----------------|----|-----------------|-------------|---------|--------|
| Model           | 54 | 581.226447704   | 10.76345328 | 129.51  | 0.0001 |
| Error           | 80 | 6.64873333      | 0.08310917  |         |        |
| Corrected Total | 134| 587.87521037    |             |         |        |
| R-Square        |    | 0.988690        |             |         |        |
| C.V             |    | 7.0817912       | 0.28828661  | 4.07081481 |
| Root MSE        |    |                 |             |         |        |
| SEEDY Mean      |    |                 |             |         |        |

| Source            | DF | Anova SS | Mean Square | F Value | Pr > F |
|-------------------|----|----------|-------------|---------|--------|
| BLOCK             | 2  | 0.3543570| 0.1771785   | 2.13    | 0.1253 |
| SHLEVEL           | 4  | 450.8827659| 112.7206915| 1356.30 | 0.0001 |
| BLOCK*SHLEVEL     | 8  | 0.8687763 | 0.1085970   | 1.31    | 0.2522 |
| VAR               | 8  | 94.0237570| 11.7529696  | 141.42  | 0.0001 |
| SHLEVEL           | 32 | 35.0968207| 1.0967756   | 13.20   | 0.0001 |
A Case Study from Coconut Research Institute of Sri Lanka

Tests of Hypotheses using the Anova MS for BLOCK*SHLEVEL as an error term

| Source          | DF | Anova SS       | Mean Square | F Value | Pr — F |
|-----------------|----|----------------|-------------|---------|--------|
| BLOCK           | 2  | 0.3543570      | 0.1771785   |         |        |
| SHLEVEL         | 4  | 450.88276593   | 112.72069148| 1037.97 | 0.0001 |

Analysis 3. (RBD with unequally replicated control)

In this experiment the control treatment was replicated three times while the six treatments (3 factors x 2 levels) were replicated once within a block. This was a long term experiment and in such cases post-treatment yield is often analysed with a pre-treatment yield as a covariate. The adjusted ANOVA table prepared manually is shown in Fig. 7.

Fig. 7 Manual analysis for the data set 3

| Source               | DF | SS    | MS    | F Value |
|----------------------|----|-------|-------|---------|
| Block                | ...| 51,532| 25,766| 6.14    |
| Control Vs. Sources | ...| 64,005| 64,005| 15.24   |
| Between levels       | ...| 9,315 | 9,315 | 2.22    |
| Between Sources      | ...| 1,549 | 775   | .18     |
| Source x level       | ...| 13,745| 6,872 | 1.64    |
| Residual             | ...| 71,391| 4,199 |         |

Adjusted means

Control
Source 1
Source 2
Source 3

SE for difference between sources = 39.7 (17 d f)
SE for difference between and source = 36.2 (17 d f)

The analyses of covariance cannot be carried with PROC ANOVA. Hence the GLM procedure was used. The SAS program to produce similar results as in Fig. 7 is shown in Fig. 8. The SAS program is relatively straightforward but not as simple as in the first two examples. The results for PROC GLM is given in Fig. 9.

Fig. 8 SAS program for data set 3

LIBNAME MAN '/SARATH/';
DATA MAN. DATA;
INPUT LEVEL SOURCE BLOCK x Y;
CARDS:

|   |   |   |   |
|---|---|---|---|
| 0 | 1 | 1 | .982 807 |
| 0 | 1 | 2 | 1068 669 |
| 0 | 1 | 3 | 655 674 |
| 0 | 2 | 1 | 838 579 |
| 0 | 2 | 2 | 922 749 |
| 0 | 2 | 3 | 569 558 |
| 0 | 3 | 1 | 947 719 |
| 0 | 3 | 2 | 474 546 |
| 0 | 3 | 3 | 219 512 |
| 1 | 1 | 1 | 791 766 |
| 1 | 1 | 2 | 992 876 |
| 1 | 1 | 3 | 463 636 |
| 1 | 2 | 1 | 878 730 |
| 1 | 2 | 2 | 555 700 |
| 1 | 2 | 3 | 310 589 |
| 1 | 3 | 1 | 828 691 |
| 2 | 3 | 2 | 467 616 |
| 1 | 3 | 3 | 492 685 |
| 2 | 1 | 1 | 938 709 |
| 2 | 1 | 2 | 762 818 |
| 2 | 1 | 3 | 447 656 |
| 2 | 2 | 1 | 840 729 |
| 2 | 2 | 2 | 578 854 |
| 2 | 2 | 3 | 617 658 |
| 2 | 3 | 1 | 785 790 |
| 2 | 3 | 2 | 523 760 |
| 2 | 3 | 3 | 383 754 |

RUN;
DATA MAN; SET MAN. DATA;
IF LEVEL=0 THEN L=0; IF 0<LEVEL<=2 THEN L=1;
IF LEVEL=0 AND SOURCE<4 THEN SOR=0; ELSE SOR=SOURCE;
PROC GLM;
CLASS BLOCK SOR L LEVEL;
MODEL Y=BLOCK X L SOR LEVEL SOR*LEVEL/SSID;
LSMEANS SOR LEVEL SOR*LEVEL/STDERR; RUN;

Fig. 9 SAS output for the data set 3.

General Linear Models Procedure

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|--------|----|----------------|-------------|---------|--------|
| Model  | 9  | 155,029.43258  | 17,225.49251 | 4.10    | 0.0060 |
| Error  | 17 | 71,391.08593   | 4,199.47564  |         |        |
| Corrected Total | 26  | 226,420.51852 |             |         |        |

R-Square | C.V. | Root MSE | Y Mean
0.684697 | 9.2920380 | 64.803361 | 697.40740741

48
A Case Study from Coconut Research Institute of Sri Lanka

Department Variable: Y

| Source     | DF  | Type I SS | Mean Square | F Value | Pr > F |
|------------|-----|-----------|-------------|---------|--------|
| BLOCK      | 2   | 51,532.741| 25,766.370  | 6.14    | 0.0099 |
| X          | 1   | 14,882.288| 14,882.288  | 3.54    | 0.0770 |
| L          | 1   | 64,005.033| 64,005.033  | 15.24   | 0.0011 |
| SOR        | 2   | 1,549.327 | 774.664     | 0.18    | 0.8332 |
| LEVEL      | 1   | 9,315.484 | 9,315.484   | 2.22    | 0.1547 |
| SOR*LEVEL  | 2   | 13,744.560| 6,872.280   | 1.64    | 0.2239 |

General Linear Models Procedure

Least Squares Means

| SOR | Y    |
|-----|------|
| 0   | Non-est |
| 1   | Non-est |
| 2   | Non-est |
| 3   | Non-est |

General Linear Models Procedure

Least Squares Means

| LEVEL  | Y    |
|--------|------|
| 0      | Non-est |
| 1      | Non-est |
| 2      | Non-est |

General Linear Models Procedure

Least Squares Means

| SOR | LEVEL | Y    |
|-----|-------|------|
| 0   | 0     | Non-est |
| 1   | 1     | Non-est |
| 1   | 2     | Non-est |
| 2   | 1     | Non-est |
| 2   | 2     | Non-est |
| 3   | 1     | Non-est |
| 3   | 2     | Non-est |
T. S. G. PEIRIS and R. D. STERN

It can easily be seen that this output is more complex than for the Split plot analysis. Different values of sums of square (Type I and Type III) are introduced for the same factors. Consequently different values of mean sums of squares and F values are produced. In this analysis Type I SS is for the SS of unadjusted ANOVA table which is an additional output than requested. The individual adjusted values cannot be obtained from SAS. Further there is no facility to test the adjusted means from SAS.

Analysis 4 repeated measurements experiment

In this experiment the total number of coconut fronds and those that were attacked by the rhinoceros beetle were counted in 64 palms (8 palms with each of 8 treatments) at six equal intervals including the pre-treatment stage. Treatments were applied once. This is a typical repeated measurements set of data which can be analysed in a variety of ways including split plot analysis, multivariate analysis and ante-dependence covariance analysis (Kenward, 1987). The manual analysis for the experimenter consisted simply of five separate 2-way ANOVA tables. The variable considered was the difference between the percentage of fronds attacked at consecutive periods. This can easily be obtained from SAS. In addition, other methods of analysis can easily be conducted. Thus, this set of data is typical of many that are already partially analysed manually and require a comprehensive statistical package for there to be any chance of realising the full potential of the data. In addition to using a computer package for analyses that are impracticable by hand, it is also valuable to use computer to present simple graphical displays. This is often of much benefit at the start of an analysis and is very tedious to do manually. Plots of two of the treatments are given in Fig. 10. We feel such plots can be of considerable benefit to the experimenter.

DISCUSSION

The statistical package SAS is an extremely powerful tool for the analysis of the experimental data collected by the Coconut Research Institute. The example in the last section shows that the presentation of results for those analyses that were done efficiently prior to computerization is not always ideal. Some notes will have to be supplied with the results and the output could be edited to bring it closer to the form that is needed by the experimenters. An alternative would be to use a different statistical package, for some of the standard analysis instead. The use of GENSTAT Version 5 is investigated in a future study, though there are facilities in SAS, such as high resolution plotting that are not available in GENSTAT. When staff become experienced in using a statistical package, the production of the standard analyses should be far quicker than previously. It will be important to utilise the time saved constructively to do more analyses than were possible manually. These should include more work on the simple analysis and presentation of data. Some experiments may require a range of more complex analyses. Perhaps, the most important aspect will be to encourage experimenter to bring the raw data (rather than the semi-processed data) to the Biometry Unit. The repeated copying...
Fig. 10. % of fronds attacked in treatment 1 in eight palms
and summary of the observations can lead to many errors as well as loss of information. The computers could be used for the storage of both the raw data and information about the experiments and not just as a large statistical calculator. This requires a well defined system of data entry and management. Although SAS could be used for this purpose, a standard database package such as DATAEASE may be preferable.

The observations reported here indicate that the computers and SAS provide tremendous potential for management and analysis of data at the CRI. Similar advantages are likely at other agricultural research institutes. It will however require careful planning, as well as hard work, to ensure that this potential is realised.

ACKNOWLEDGEMENTS

The authors thank the Heads of Divisions of Soils and Plant Nutrition, Crop Protection and Agronomy of the Coconut Research Institute for allowing the use of their experimental data. They are also thankful to the Computer and Information Technology Council (CINTEC) of Sri Lanka and the British Council in Colombo for providing a fellowship to TSGP, which enabled this study to be carried out at University of Reading. The authors also wish to thank Dr. R. Mahindapala, Director, CRI, for his valuable comments.

REFERENCES

DATAEASE Version 2.5 Quick Reference Guide (1986): Sapphire Systems Ltd., UK.

Kenward, M. G. (1987): A method for comparing profiles of repeated measurements. *J. App. Stats.* 36 (3): 296-308.

SAS/STAT guide for Personal Computers (1986), SAS Institute Inc., USA.