ANALYSIS OF AUTOCORRELATED DATA IN GROUPS OF EXPERIMENTS

By

PREMI T. C.

THESIS

Submitted in partial fulfilment of the
requirements for the degree of

Master of Science in Agricultural Statistics

Faculty of Agriculture
Kerala Agricultural University

Department of Agricultural Statistics
COLLEGE OF HORTICULTURE
Vellanikkara - Thrissur
Kerala, India

1994

DECLARATION

I hereby declare that this thesis entitled **"Analysis of Autocorrelated Data in Groups of Experiments"** is a bonafide record of research work done by me during the course of research and that the thesis has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or any other similar title, of any other University or Society.

Vellanikkara
23.11.94

PREMI, T.C.

CERTIFICATE

Certified that this thesis entitled **"Analysis of Autocorrelated Data in Groups of Experiments"** is a record of research work done independently by **Mrs. Premi, T.C.**, under my guidance and supervision and that it has not previously formed the basis for the award of any degree, fellowship or associateship to her.

Vellanikkara
23-11-94

V.K. Gopinathan Unnithan
(Chairman, Advisory Committee)
Associate Professor of
Agricultural Statistics
College of Horticulture

CERTIFICATE

We, the undersigned members of the Advisory Committee of **Mrs. Premi, T.C.**, a candidate for the degree of Master of Science in Agricultural Statistics, agree that the thesis entitled **"Analysis of Autocorrelated Data in Groups of Experiments"** may be submitted by Mrs. Premi, T.C. in partial fulfilment of the requirement for the degree.

Shri. V.K. Gopinathan Unnithan
Associate Professor
Department of Agricultural Statistics
College of Horticulture
Vellanikkara
(Chairman)

Prof. P.V. Prabhakaran
Professor and Head
Department of Agricultural Statistics
College of Horticulture
Vellanikkara
(Member)

Smt. K.A. Mercy
Assistant Professor
Department of Agricultural Statistics
College of Horticulture
Vellanikkara
(Member)

Dr. G. Sreekandan Nair
Professor & Head
Department of Horticulture
College of Agriculture
Vellayani
(Member)

External Examiner

ACKNOWLEDGEMENTS

I express my deep sense of gratitude and indebtedness to Sri. V.K. Gopinathan Unnithan, Associate Professor, Department of Agricultural Statistics, College of Horticulture, Kerala Agricultural University, Vellanikkara and chairman of the advisory committee for suggesting the problem, excellent guidance and overwhelming inspiration throughout the course of the investigation and preparation of the thesis.

My sincere thanks to are due to Prof. P.V. Prabhakaran, Professor & Head, Department of Agricultural Statistics, College of Horticulture, Vellanikkara, Smt. K.A. Mercy, Assistant Professor, Department of Agricultural Statistics, College of Horticulture, Vellanikkara, Dr. G. Sreekandan Nair, Professor & Head, Department of Horticulture, College of Agriculture, Vellayani, Members of the Advisory Committee for the pains taking involvement throughout the course of the programme.

I would like to express my sincere thanks to the Associate Dean, College of Horticulture and Dean, College of Veterinary and Animal Sciences for providing necessary facilities for the study.

I will be failing my duty if I do not express my extreme gratitude to the staff members of the Department of Statistics and my fellow students for their co-operation and sincere interest in my work.

I am grateful to the Kerala Agricultural University for the award of junior fellowship during the course of study.

I am highly obliged to all my friends and well-wishers for their assistance and kind co-operation.

My sincere thanks to Smt. Geetha Bai, S. and Sri. O.K. Ravindran, C/o Peagles, Mannuthy for the neat typing of the manuscript.

My sincere thanks to professor & head of the Cadbury-KAU Co-operative Cocoa Research Project, College of Horticulture, Kerala Agricultural University for providing the necessary facilities for collection of experimental data.

Mere gratitude is not enough to express the constant encouragement and affection rendered by my family members at all stages of the investigation.

Above all, I bow my head before the Almighty God who blessed me with health and confidence to undertake the work successfully.

PREMI, T.C.

CONTENTS

LIST OF TABLES

Introduction

INTRODUCTION

Agricultural experiments carried out at Research Stations are to formulate recommendations on farm practices for a population quite extensive in space or time or both. Therefore it becomes necessary to ensure that the results obtained from such experiments are valid for several seasons and over a reasonably heterogeneous space. A single experiment furnishes information about the tract where it is conducted. Hence it is a common practice to repeat an experiment at different places or over a number of occasions or both to provide valid recommendations.

Observations are recorded repeatedly at selected intervals of time, in experiments on animals, human beings and perennial crops. The analysis of data so generated are complicated by the need to consider two sources of variation, viz., the usual 'among subject' variation and the 'within-subject' variation. The presence of serial correlation among the error terms violates the assumption of independence of error terms for the analysis of variance. Another difficulty is that simple analysis of variance involving years are useful only if we can assume that there are no residual effects of crops or treatments as the case may be.

Statistical methods for comparing means using the t and F distributions are valid only if the error terms are independently as well as normally distributed. Standard analytical procedures can be justified in the sense that they yield valid sampling distribution, if proper randomization is done while conducting the experiment. But in experiments with repeated measurements randomization remains unchanged, with regard to successive observations on the same individual or experimental unit and hence the error terms can no longer be considered independent.

In recent years there have been increasing activity in the development of procedures in this problem. The present practice is to analyse the data in a split-plot set up. But the presence of serial correlation among the error terms violates the assumption of constant correlation between time points within a unit. Sukumaran (1991) proposed a model for situations where repeated observations are taken on the same set of experimental units, taking the dependence of error terms into consideration. But the estimation of parameters could not be developed satisfactorily. Hence the present investigation was taken up to make suitable modification in the model proposed by Sukumaran (1991) and to develop procedure of estimation of parameters and comparison of treatment means.

Review of Literature

REVIEW OF LITERATURE

In experiments with repeated measurements, the measurements have a temporal sequence with the consequence that measurements on the same subject separated in small time intervals will in general be highly correlated. various procedures that have been suggested or used by different workers for the analysis of groups of experiments with special reference to data on repeated measurements have been reviewed. They are presented briefly in this chapter in various sections viz., general, split-plot analysis, ARMA model, analysis of differences and linear model.

Yates and Cochran (1938) proposed the analysis of data from a set of experiments involving same or similar treatments carried out at a number of places or in a number of years or both. They pointed out that the ordinary analysis of variance procedure suitable for dealing with the results of a single experiment may require modification, owing to lack of equality in the errors of the different experiments and owing to non-homogeneity in the components of the interaction of treatments with places or times or both.

Wallenstein (1982) criticised the use of standard analysis of variance for experiments with repeated observations and observed that the analysis of Aitken (1981) was based on assumptions that could not hold under such situations.

Rowell and Walters (1976) criticised the split-plot analysis of experimental data where several successive observations of the same variable have been recorded on each experimental unit on the ground that requirements for such analysis received scant attention and it was often unlikely that these assumptions would be satisfied in experimental situations. They presented five sets of results to support this proposition. They also proposed an alternative analytical approach in which contrasts over time are analysed.

Multivariate analysis

Cole and Grizzle (1966) proposed multivariate analysis of variance thereby eliminating the problems of univariate analysis variance model. According to them this procedure provides ab unified approach to the analysis of such data with all the power, scope and flexibility of the univariate analysis variance.

Yates (1954) investigated the problems arising in the analysis of experiments containing different crop rotations. He proposed analysis of data by fitting constants to overcome certain difficulties in such solutions.

Khosla et al (1979) investigated the behaviour of experimental errors and presence of treatments x years interactions in the case of 199 groups of experiments conducted at different research stations in the state of Gujarat during different years of the period 1960-65.

Hearne et al (1983) reported that a popular design in biostatistics and psychometrics is the time factor experiment involving one grouping factor and one repeated measures factor. If the F-tests involving the repeated measures factor are to be valid, univariate analytic procedures require special assumptions for covariances.

Raz (1989) reported that the mixed model analysis of variance, that is commonly applied to repeated measurements taken over time, depends on specialized assumptions about the error distribution and fails to exploit information contained in the ordering of the data points over time. He developed a procedure that overcomes these disadvantages while preserving familiar features of the mixed model ANOVA. Group profiles are estimated by non-parametric smoothing of observed mean profiles. Group and time main effects, and the group by time interaction effect, are tested using randomization tests. He proposed a new approximate F-test for time effect.

Gaylor (1978) demonstrated the effects from unrecognized correlations or due to ignoring correlations among errors upon the analysis of variance. He concluded that the assumption of uncorrelated errors is not always fulfilled for an analysis of variance and this would negate the exactness of F-tests.

Jennrich and Schluchter (1986) conducted analysis of incomplete repeated-measures data by maximum likelihood method using a general linear model for expected responses and arbitrary structural models for the within subject covariances.

Split-plot analysis

Patternson (1939) considered the problem of field experimentation with perennial crops and suggested that certain modifications have to be effected in statistical analysis of long term data on perennial crops. He recommended the use of split plot design for the analysis of long term experiments with years assigned to subplots and treatments assigned to main plots.

Steel et al (1960) supported the split-plot design approach for analysis of experiments in which repeated observations are made.

Yates (1982) opinioned that the split-plot analysis suggested by Aitkin (1981) for an experiment in which rectal temperature were measured was incorrect. He found that the six values for any one subject exposed to a particular ambient temperature can be fitted exactly by a fifth degree polynomial. He observed that the analysis using orthogonal polynomials was conceptually and computationally simple though it may not provide the best approach in specific problems.

Aitkin (1981) found that regression models could be used for response on subjects measured repeatedly under different experimental conditions and he called such designs as split-plot designs. He observed that depending on the way the treatment design is set up, some effects might be tested against within subject variation and others against among subject variations. He illustrated the procedure using data generated from an experiment on animals. He suggested maximum likelihood procedure in presence of serial correlation.

Kenward (1984) proposed ANOVA for repeated measurements in which the errors have autoregressive covariance structure. By simulation he demonstrated that the distribution of the resulting trace statistic is well approximated by F distribution.

Gill (1986) reported that animal scientists who conducted experiments involving repeated measurements of animals often are frustrated by low statistical power of tests for comparisons of treatment means. In many cases, low power of the traditional tests simply is the consequence of low replication (few animals per treatment) that was forced by cost or complexity of experimental technique. A method is proposed for comparing treatments in a way that permits sensitive tests when the number of animals per treatment is not more than five or six.

Dempster (1963) extended the stepwise testing of multivariate analysis of variance to linear combination of variables resulting from a principal component analysis.

Danford et al. (1960) opinioned that the assumptions and techniques of the usual univariate analysis of variance procedure for data involving repeated measurements on the same individuals over time are not necessary and proposed multivariate procedures.

Steel (1955) proposed a bivariate analysis of variance for perennial crop data. He made a test of hypotheses about varietal effects and compared the bivariate analysis and a univariate analysis.

Smith *et al*. (1962) illustrated, using data taken from a study made by Brown and Beerstecher (1951), some methods of analysis and interpretation under the model I of MANOVA. Three criteria for the tests of significance were illustrated. They gave a rather informal way of approaching the multiple comparisons problem.

Evans and Roberts (1979) proposed fitting polynomial equations to the sequences and analysing their coefficients by MANOVA as they are correlated between experimental units.

Finney (1956) severely criticised the use of multivariate analysis of variance, construction of cannonical variates for the analysis for interpretation of agricultural experiments repeated over years. He observed that the method described by Steel (1955) to be tedious and pointed out that Steel (1955) did not make the aim of experiment clear and that the manner in which his analysis enabled the agricultural scientists to reach conclusion relevant to the problem could have better been obtained by simple alternatives.

ARMA models

Bjornson (1978) proposed an autoregressive moving average model ARMA (p,q) for the error terms in perennial crop experiments, which are autocorrelated. The problem of identifying and estimating the ARMA (p,q) error models for experimental error differs in some aspects from the common time series applications, and an error component independent of time differences, plot error may also be present. The errors in a group of experiments were identified as second order autoregressive AR(2). A transformation was suggested for the data with AR(2) residuals prior to the regression analysis of the vectors of annual treatment responses.

Yand and Randy (1983) conducted an experiment in which successive observations over time made on individual subjects were classified into different groups. They observed that successive observations on each individual follow a simple ARMA model and concluded that in many practical situations the usual analysis of variance F test, performed on averages of the observations over time provides an efficient test.

Laing and Panhuber (1988) observed that in a two-way analysis of variance, one factor may have a spatial or temporal structure, in which case correlation may be expected between the errors for the different levels of the factor. They developed a procedure for checking this using correlation coefficients.

Analysis of increments

Box (1950) concluded that the effects can often be simply interpreted by differencing the original data, in the analysis of growth and wear curves. These differences correspond to the average growth rates during successive periods. Successive periods are treated as the levels of a further factor-periods, the effect of treatments on mean rate is measured by the variation in the period averages and on the shape of the rate curve by the interaction of these treatments with periods. The taking of differences sometimes results in a very simple covariances pattern for the errors, and the analysis can be made by the technique of analysis of variance. A test is given which makes it possible to decide whether this simple set up is contradicted by the data. If it is not appropriate, a multivariate extension of the analysis of variance is suggested to make the tests. Certain simple properties of the criterion are discussed which facilitate the analysis and the elimination of variables such as initial weight. Finally an important assumption is made that the variables are multinormally distributed about their mean values with constant variance and covariance matrix in the multivariate analysis.

Gill (1988) proposed that reduction of inter-period correlation by using first differences does not necessarily eliminate problems with heterogeneity of the variance matrix over time. But adhoc claims (Box, 1950) that analysis of increments of response was superior to trend analysis of the original data with a split-plot model as spurious. For homogeneous condition the expected variance of a simple trend contrast between two treatments for adjacent periods is shown to be the same for either analyses, but the analysis of increments incurs a loss of degrees of freedom that can be critical in studies with few experimental units per treatments.

Linear models

Sukumaran (1991) proposed a model for the analysis of repeated measurements from the same experimental unit and is given as

$$Y_{ijk} = \mu + t_i + b_j + p_k + t_{ik} + b_{jk} + \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1} + e_{ijk}$$

where μ is the overall effect, t_i, the effect of the i^{th} treatment, b_j, the effect of the j^{th} replication, p_k, the effect of the k^{th} year, t_{ik}, the interaction effect of the i^{th} treatment and k^{th} year, b_{jk}, the interaction effect of j^{th} replication and k^{th} year, β_{n-1}^{k} the partial regression coefficient of $(Y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk})$ on e_{ijk-1} and e_{ijk} the error term attached to the observation Y_{ijk} which are assumed independent among themselves and with other terms in the model, and normally distributed with mean zero and constant variance say σ_e^2. The dependence of the error terms of successive observations was incorporated in the proposed model. The error mean square using the model was derived using the principle of least squares. The proposed methodology was also illustrated.

Geary (1989) proposed a procedure for sequential testing in the context of repeated significance tests in clinical traits, in which repeated measurements of the same variable are made on a fixed number of subjects over a period of time.

Pantula and Pollock (1985) proposed an auto regressive time series model for a randomized experimental design with several successive time measurements on each experimental unit. The linear model for individuals selected at random and measurements taken at t_1 consecutive time points on the i^{th} individual was given by

$$Y_{ij} = \sum_{k=1}^{r} G_{ijk} \phi_k + a_{ij}$$

where Y_{ij} denotes the value of the j^{th} measurement for the i^{th} individual, G_{ijk}, $k=1$ ..... r denote the levels of the r control variables at which the observations Y_{ij} is obtained, ϕ_k denotes the unknown parameters to be estimated and a_{ij}, the random error associated with y_{ij}.

Mansour et al(1985) proposed linear models used for the estimation of variance components formulated under the assumption of independent errors. The model was given by

$$y_{ij} = \mu + a_i + T_j + e_{ij}$$

where y_{ij} is the score recorded for the i^{th} subject at the j^{th} time, μ the overall fixed mean, a_i the random effect of i^{th} subject and $a_i \sim N(O, \quad a^2)$, T_j the fixed effect of j^{th} time and e_{ij} the random error term. It is a two way mixed model with errors assumed to follow a first order, non-stationary autoregressive process. Maximum likelihood techniques were recommended to estimate the variance components and parameters for the autoregression.

Diggle (1988) proposed a linear model for repeated measurements in which correlation structure within each time sequence of measurements included parameters for measurement error, variation between experimental units, and serial correlation within units. He developed an approach to data analysis which involved preliminary analysis by ordinary least squares, use of empirical semivariogram or residuals to suggest a suitable correlation structure and formal inference using likelihood based methods.

Cullis and McGilchrist (1990) developed stochastic differential equation of Sandland and McGilchrist (1979) and presented a growth model from designed repeated measurements. Residual maximum likelihood (REML) method was used to estimate the parameters in the model. The model was also extended to incomplete data to overcome some of the practical difficulties encountered with the profile model.

Methodology

METHODOLOGY

Pooled analysis of data generated from multi-locational trials have been developed and is in wide use satisfactorily. But the available procedure can't be used directly when the error terms in the analysis of variance model are not independent. The error terms in the analysis of variance model of experiments generating information by repeated measurements on the same experimental unit are dependent.

We consider a comprehensive model incorporating the possible relationship of the residual terms. Without loss of generality, let us consider an experiment involving 't' treatments, replicated 'r' times, laid out in RBD and observations taken on q years (occasions) on the same experimental units.

Let the observation from the experimental unit in the j^{th} replication receiving i^{th} treatment at the k^{th} year , be y_{ijk}. The model proposed by Sukumaran (1991) is

$$y_{ijk} = \mu + t_i + b_j + p_k + t_{ik} + b_{jk} + \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1} + e_{ijk} \qquad \text{------ (1)}$$

where μ is the overall effect

t_i - the effect of the i^{th} treatment

b_j - the effect of j^{th} replication

p_k - the effect of the k^{th} year

t_{ik} - the interaction effect of the i^{th} treatment and k^{th} year

b_{jk} - the interaction effect of j^{th} replication and k^{th} year,

β^{k}_{n-1} - the partial regression coefficient of

$(y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk})$ on e_{ijn-1} and e_{ijk} the error term attached to the observation y_{ijk} which are assumed independent among themselves and with other terms in the model, normally distributed with mean zero and constant variance say σ_e^2.

Estimation of parameters and various sum of squares in this model is very complicated. Hence modification on this model is proposed here in to simplify the whole approach. The error term attached to the observation on an experimental unit of any year is assumed to depend only on that of the immediately preceding year. Thus the model can be written as,

$$y_{ijk} = \mu + t_i + b_j + p_k + t_{ik} + b_{jk} + \phi_k\, e_{ijk-1} + e_{ijk} \qquad \text{------ (2)}$$

Where ϕ_k is the regression coefficient of on $(y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk})$ on e_{ijk-1}

The parameters in the model can be estimated by the method of least squares.

The error sum of square is given by

$$R = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1})^2 \quad \text{------- (3)}$$

The summation being taken over all values of i, j and k.

Estimates of the parameters μ, t_i, b_j, p_k, t_{ik}, b_{jk} and ϕ_k's are obtained by solving the set of normal equations derived by setting the first differential of R with respect to each of these parameters to zero. Thus differentiating R with respect to μ, t_i, $i = 1, 2, \ldots\ldots t$, b_j, $j = 1, 2 \ldots\ldots r$, p_k, $k = 1 \ldots\ldots q$, t_{ik}, b_{jk} and ϕ_k and equating to zero, we get a set of normal equations. But the number of independent normal equations obtained is less than the number of parameters to be estimated. Therefore we impose the following restrictions as usual.

$$\sum_{i=1}^{t} t_i = \sum_{j=1}^{r} b_j = \sum_{k=1}^{q} p_k = \sum_{i=1}^{t} t_{ik} = \sum_{k=1}^{q} t_{ik} = \sum_{j=1}^{r} b_{jk} = \sum_{k=1}^{q} b_{jk} = 0 \quad \text{------ (4)}$$

As a consequence $\sum_{ij} e_{ijk} = 0$, $k = 1, 2, \ldots\ldots q$

$$\sum_{i=1}^{r} e_{ijk} = \sum_{j=1}^{r} e_{ijk} = 0 \quad i = 1, 2, \ldots\ldots, t,$$
$$j = 1, \ldots\ldots, r,$$
$$K = 1, 2, \ldots\ldots q \quad \text{------} \ (5)$$

Now,

Differentiating R' with respect to μ , and equating to zero and imposing restriction like (4) we get.

$$-2 \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1}) = 0$$

This yields $\hat{\mu} = \frac{y_{\ldots}}{qrt}$, where a dot replacing a subscript indicates summation with respect to that subscript.

Differentiating R' with respect to t_i, and equating to zero and imposing restriction like (4) we get.

$$-2 \sum_{jk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1}) = 0$$

i.e. $\sum_{jk} y_{ijk} - \sum_{jk} \quad - \sum_{jk} t_i = 0, \ i = 1,2, \ldots\ldots t$

Therefore $\hat{t}_i = \frac{y_{i..}}{qr} - \frac{y_{\ldots}}{qrt}, \ i = 1,2, \ldots\ldots t$

Differentiating R' with respect to b_j, and equating to zero and imposing restriction like (4) we get.

i.e., $$-2 \sum_{ik} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1}) = 0$$

i.e. $$\sum_{ik} y_{ijk} - \sum_{ik} \mu - \sum_{ik} b_j = 0$$

Therefore $$\hat{b}_j = \frac{y_{.j.}}{qt} - \frac{y_{...}}{qrt} , \; j = 1, \; .... \; r$$

Differentiating R with respect to p_k, and equating to zero and imposing restriction like (4) we get.

$$-2 \sum_{ij} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1}) = 0$$

i.e. $$\sum_{ij} y_{ijk} - \sum_{ij} \quad - \sum_{ij} p_k = 0$$

Therefore $$\hat{p}_k = \frac{y_{..k}}{rt} - \frac{y_{...}}{qrt} , \; k = 1,2, \; ...... \; q$$

Differentiating R with respect to t_{ik} , and equating to zero and imposing restriction like (4) we get

$$\therefore \quad \sum_{j} y_{ijk} - \sum_{j} \mu - \sum_{j} t_i - \sum_{j} p_k - \sum_{j} t_{ik} = 0$$

i.e. $$\hat{t}_{ik} = \frac{y_{i.k}}{r} - \hat{\mu} - \hat{t}_i - \hat{p}_k$$

$$= \frac{y_{i.k}}{r} - \frac{y_{i..}}{qr} - \frac{y_{..k}}{rk} + \frac{y_{...}}{qrt} , \; i = 1,2 \; .... \; t,$$

$$k = 1,2 \; ....q$$

Differentiating R' with respect to b_{jk}, and equating to zero and imposing restriction like (4) we get.

$$-2 \sum_{i} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1}) = 0$$

i.e. $\sum_{i} y_{ijk} - \sum_{i} \mu - \sum_{i} b_j - \sum_{i} p_k - \sum_{i} b_{jk} = 0$

Therefore $\hat{b}_{jk} = \frac{y_{.jk}}{t} - \frac{y_{.j.}}{qt} - \frac{y_{..k}}{rt} + \frac{y_{...}}{qrt}$

$j = 1,2 \ldots\ldots r, \quad k = 1,2 \ldots\ldots q$

Differentiating R' with respect to ϕ_k, and equating to zero and imposing restriction like (4) we get

$$-2 \sum_{ij} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1}) e_{ijk-1} = 0$$

i.e.,

$$\sum_{ij} y_{ijk} e_{ijk-1} - \mu \sum_{ij} e_{ijk-1} - \sum_{ij} t_i e_{ijk-1} - \sum_{ij} b_j e_{ijk-1}$$

$$- \sum_{ij} p_k e_{ijk-1} - \sum_{ij} t_i e_{ijk-1} - \sum_{ij} b_{jk} e_{ijk-1}$$

$$- \sum_{ij} \phi k \, e^2_{ijk-1} = 0, \; k = 2,3 \ldots\ldots q$$

i.e. $\sum_{ij} y_{ijk}\, e_{ijk-1} - \sum_{i} t_i\, e_{i.k-1} - \sum_{j} b_j\, e_{.jk-1} -$

$\sum_{ij} p_k\, e_{..k-1} - \sum_{i} t_{ik}\, e_{i.k-1} - \sum_{j} b_{jk}\, e_{.jk-1} -$

$\sum_{ij} \phi k\, e^2_{ijk-1} = 0$

i.e. $\sum_{ij} y_{ijk}\, e_{ijk-1} - \phi k \sum_{ij} e^2_{ijk-1} = 0, \; k = 2,3 \ldots. \; q$

Therefore $$\hat{\phi} k = \frac{\sum_{ij} y_{ijk}\, e_{ijk-1}}{\sum_{ij} e^2_{ijk-1}} \qquad \text{------- (6)}$$

ϕk's depend on the observations on the k^{th} year and the error terms of the previous year. ϕk's are to be estimated for the values of k ranging from 2 to q. $\phi 2$ can be estimated using (6) from the observations of the year 2 and the residual term of the first year (e_{ij1}'s).

Once the estimate of ϕ 2 is thus obtained, the residual term (error term) of the second year, e_{ij2}'s can be obtained. From these residual terms of the second year and the observed values of the third year, ϕ 3 can very well be estimated using (6). Proceeding like this all ϕ k values and thus all parameters of the model (2) can be estimated.

To find error sum of squares

Substituting the estimators of parameters in eq. (3) we get

$$R = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1})^2$$

$$= \sum_{ijk} y_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \phi_k e_{ijk-1})$$

$$= \sum_{ijk} y^2_{ijk} - \hat{\mu}\, y_{...} - \sum_i y_{i..}\, \hat{t}_i - \sum_j y_{.j.}\, \hat{b}_j - \sum_k y_{..k}\, \hat{p}_k -$$

$$\sum_{ik} \hat{t}_{ik}\, y_{i.k} - \sum_{jk} \hat{b}_{jk}\, y_{.jk} - \sum_{ijk} \hat{\phi}_k\, e_{ijk-1}\, y_{ijk}$$

$$= \sum_{ijk} y^2_{ijk} - \sum_k \frac{y^2_{i.k}}{r} - \sum_{jk} \frac{y^2_{.jk}}{t} + \sum_k \frac{y^2_{..k}}{rt} -$$

$$\sum_k \phi_k \sum_{ij} y_{ijk}\, e_{ijk-1} \qquad \text{-------} \quad (7)$$

In order to derive the expression for the sum of squares due to the effect of each factor in equation (2) a hypothesis of null differences among the different levels of each factor which amounts to zero effect for each level of each factor due to the restriction of the zero total effect is made. The resultant residual sum of squares say R', contains the sum of squares due to the effect of that factor on which the hypothesis is made. Therefore R' - R will provide the sum of squares due to that factor. In a similar manner S.S. due to every factor can be derived.

To estimate treatment sum of square and testing its significance

$$H_o : t_i = t_2 = \ldots\ldots = t_t$$

In order to test this hypothesis H_o, a model which does not have the treatment x season interaction may be considered i.e. we consider model

$$y_{ijk} = \mu + t_i + b_j + p_k + b_{jk} + \phi'_k e'_{ijk-1} + e'_{ijk} \quad ----(8)$$

The various effects from this model can be estimated using least square technique.

The residual sum of squares under model (8) is given by

$$R^{(1)} = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1})^2 \quad ------ (9)$$

Differentiating R' with respect to μ, and equating to zero and imposing restriction like (4) we get.

$$= -2 \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1})^2 = 0$$

Therefore $\hat{\mu} = \frac{y_{...}}{qrt}$; since it can very well be shown that estimates of $e'_{..k}$ is zero, for every k.

Differentiating $R^{(1)}$ with respect to t_i and equating to zero we get

$$-2 \sum_{jk} (Y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k \, e'_{ijk-1}) = 0$$

$i = 1, 2, \ldots\ldots. t.$

i.e., $$\sum_{jk} y_{ijk} = \sum_{jk} \mu + \sum_{jk} t_i + \sum_{jk} \phi'_k \, e'_{ijk-1}$$

Therefore $$\hat{t}_i = \frac{y_{i..}}{rq} - \frac{y_{...}}{rqt} - \sum_{jk} \phi'_k \frac{e'_{ijk-1}}{rq}$$

$$= \frac{y_{i..}}{rq} - \frac{y_{...}}{rqt} - \sum_{k} \phi'_k \frac{e'_{i.k-1}}{rq}$$

Differentiating $R^{(1)}$ with respect to b_j and equting to zero, we get

$$-2 \sum_{ik} (Y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k \, e'_{ijk-1}) = 0$$

i.e., $$\sum_{ik} y_{ijk} - \sum_{ik} \mu - \sum_{ik} b_j - \sum_{ik} \phi'_k \, e'_{ijk-1} = 0$$

Therefore $\hat{b}_j = \frac{y_{.j.}}{tq} - \frac{y_{...}}{rqt}$ Since $e_{.jk-1} = 0$

Differentiating $R^{(1)}$ with respect to p_k and equating to zero we get

$$-2 \sum_{ij} (Y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1}) = 0$$

i.e., $$\sum_{ij} y_{ijk} - \sum_{ij} \mu - \sum_{ij} p_k = 0$$

Therefore $$\hat{p}_k = \frac{y_{..k}}{rt} - \frac{y_{...}}{rtq}$$

Differentiating $R^{(1)}$ with respect to b_{jk} and equating to zero, we get

$$-2 \sum_{i} (Y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1}) = 0$$

i.e., $$\sum_{i} y_{ijk} - \sum_{i} \mu - \sum_{i} b_j - \sum_{i} p_k - \sum_{i} b_{jk} - \sum_{i} \phi'_k e'_{ijk-1} = 0$$

Therefore $$\hat{b}_{jk} = \frac{y_{.jk}}{t} + \frac{y_{...}}{qrt} - \frac{y_{..k}}{rt} + \frac{y_{.j.}}{tq} +$$

$$\sum_{k} \phi'_k \frac{e'_{.jk-1}}{tq} - \phi_k \frac{e'_{.jk-1}}{t} \quad \text{Since } e'_{.jk-1} = 0$$

$$= \frac{y_{.jk}}{t} - \frac{y_{.j.}}{tq} - \frac{y_{..k}}{rt} + \frac{y_{...}}{qrt}$$

$$\frac{\partial R^{(1)}}{\partial \phi'_k} = 0$$

i.e.,
$$-2 \sum_{ij} (Y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1}) e'_{ijk-1} = 0$$

i.e.,
$$\sum_{ij} (Y_{ijk} \quad e'_{ijk-1}) - \sum_{ij} \mu e'_{ijk-1} - \sum_{ij} t_i e'_{ijk-1} - \sum_{ij} b_j e'_{ijk-1} - \sum_{ij} p_k e'_{ijk-1} - \sum_{ij} b_{jk} e'_{ijk-1} - \sum_{ij} \phi'_k e'^2_{ijk-1} = 0$$

i.e.,
$$\sum_{ij} Y_{ijk} e'_{ijk-1} - \sum_{ij} \left(\frac{y_{i..}}{rq} - \frac{y_{...}}{rqt} - \sum_k \phi'_k \frac{e'_{i.k-1}}{rq} \right) e'_{ijk-1}$$

$$- \sum_{ij} \left(\frac{y_{.j.}}{tq} - \frac{y_{...}}{rtq} \right) e'_{ijk-1}$$

$$- \sum_{ij} \left(\frac{y_{.jk}}{t} - \frac{y_{..k}}{rt} + \frac{y_{...}}{rt} - \frac{y_{.j.}}{tq} \right) e'_{ijk-1}$$

$$- \sum_{ij} \phi'_k e'^2_{ijk-1} = 0$$

$$\sum_{ij} y_{ijk} e'_{ijk-1} - \sum_{ij} \frac{y_{i..}}{rq} e'_{ijk-1} + \sum_{ij} \frac{y_{...}}{rtq} e'_{ijk-1}$$

$$+ \sum_{ij} \sum_{k} \phi_k \frac{e'_{i.k-1}}{rq} e'_{ijk-1} - \sum_{ij} \frac{y_{.j.}}{tq} e'_{ijk-1}$$

$$+ \sum_{ij} \frac{y_{...}}{rtq} e'_{ijk-1} - \sum_{ij} \frac{y_{.jk}}{t} e'_{ijk-1} + \sum_{ij} \frac{y_{..k}}{rt} e'_{ijk-1}$$

$$+ \sum_{ij} \frac{y_{.j.}}{tq} e'_{ijk-1} - \sum_{ij} \frac{y_{...}}{rt} e'_{ijk-1}$$

$$- \sum_{ij} \phi'_k e'^2_{ijk-1} = 0$$

i.e., $$\sum_{ij} e'_{ijk-1} y_{ijk-1} - \sum_{i} y_{i..} \frac{e'_{i.k-1}}{rq} + \sum_{i} \left(\sum_{k} \phi'_k e'_{i.k-1} \right) e'_{i.k-1}$$

$$- \phi'_k \sum_{ij} e'^2_{ijk-1} = 0$$

$$\hat{\phi}'_k = \frac{\sum_{ij} y_{ijk} e_{ijk-1} - \sum_{i} \frac{y_{i..} e'_{i.k-1}}{rq} + \sum_{i} \left(\sum_{k} \phi'_k e'_{i.k-1} \right) e'_{i.k-1}}{\sum_{ij} e'^2_{ijk-1}} \qquad \text{------ (10)}$$

It may be noted that the estimates of $e'_{.j.}$ is not zero even if we are imposing the restriction like (5).

It is also found that the estimates of t_i contains the term

$\sum_k \phi'_k \frac{e'_{i.k-1}}{rq}$ and estimate of ϕ'_k is a function of ϕ'_k's.

So for the estimation of ϕ'_k we propose a method of iteration. For the estimation of ϕ'_k values, e'_{ijk-1} values are needed.

On substitution of the estimates of the parameters except ϕ_k in the model (8) we get

$$y_{ijk} = \frac{y_{i..}}{rq} + \frac{y_{.jk}}{t} - \frac{y_{...}}{rqt} - \sum_k \phi'_k \frac{e'_{i.k-1}}{rq} + \phi_k e'_{ijk-1} + e'_{ijk}$$

Let $z_{ijk} = y_{ijk} - \frac{y_{i..}}{rq} - \frac{y_{.jk}}{t} + \frac{y_{...}}{rqt}$

Therefore $z_{ijk} = \phi'_k e'_{ijk-1} - \sum_k \phi'_k \frac{e'_{i.k-1}}{rq} + e'_{ijk}$

We know that, for the first year the regression coefficient is zero and thus z_{ijk}'s and e'_{ijk}'s are same

For K = 2 to q, e'_{ijk}'s can be expressed as

$$e'_{ijk} = z_{ijk} - \phi_k (e'_{ijk-1} + \sum_k \phi'_k \frac{e'_{i.k-1}}{rq}$$

Now the iterative procedure for evaluation of ϕ'_k s described below

1. Initial solution of ϕ'_k's, say $\phi'^{(o)}_k$ may be taken as the values of ϕ_ks obtained for model (2)

2. Solution of ϕ'_k's at the first iteration say $\phi'^{(1)}_k$'s can be obtained by substituting $\phi_k^{(o)}$ in the RHS of (10).

3. Solution of ϕ_k's at the second iteration say $\phi_k^{(2)}$ can be obtained by substituting $\phi_k^{(1)}$'s in the RHS of (10) and the solution of (r+1) st iteration, say $\phi_k^{(r+1)}$ are obtained by substituting $\phi_k^{(r)}$ in the RHS of (10).

4. The iteration may be continued until two successive iterations coincide upto some prespecified accuracy.

To find error sum of square

Substituting the estimates of parameters in eq (9) we get

$$R^{(1)} = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1})^2$$

$$= \sum_{ijk} y_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - b_{jk} - \phi'_k e'_{ijk-1})$$

$$= \sum_{ijk} y^2_{ijk} - \sum_{i} \frac{y^2_{i..}}{rq} - \sum_{jk} \frac{y^2_{.jk}}{t} + \frac{y^2_{...}}{rtq} +$$

$$\sum_{k} \phi_k \sum_{i} \frac{y_{...} e'_{i.k-1}}{rq} + \sum_{ijk} \phi'_k y_{ijk} e'_{ijk-1}$$

------ (11)

Treatment x season sum of square is obtained by subtracting (7) from (11).

Under H_o the model (8) reduces to

$$y_{ijk} = \mu + b_j + p_k + b_{jk} + \phi''_k \, e''_{ijk-1} + e''_{ijk} \qquad \text{------ (12)}$$

The various effects from this model can be estimated using principle of least squares. By minimising

$$R^{(2)} = \sum_{ijk} (y_{ijk} - \mu - b_j - p_k - b_{jk} - \phi''_k \, e''_{ijk-1})^2$$

The estimates are

$$\hat{\mu} = \frac{y_{...}}{qrt} \text{ ; since } e_{..k} = \text{zero, for every k.}$$

$$\frac{\partial R^{(2)}}{\partial b_j} = 0$$

i.e.,

$$-2 \sum_{ik} (y_{ijk} - \mu - b_j - p_k - b_{jk} - \phi''_k \, e''_{ijk-1}) = 0$$

i.e.,

$$y_{.j.} - tq \quad - tqb_j - \sum_{ik} \phi''_k \, e''_{.jk-1} = 0$$

Therefore

$$\hat{b}_j = \frac{y_{.j.}}{tq} - \frac{y_{..}}{rtq} - \sum_k \phi''_k \frac{e''_{.jk-1}}{tq}$$

$$\frac{\partial R^{(2)}}{\partial p_k} = 0$$

gives $\hat{p}_k = \frac{y_{..k}}{rt} - \frac{y_{...}}{rtq}$

$$\frac{\partial R^{(2)}}{\partial b_{jk}} = 0$$

i.e., $\sum_i (y_{ijk} - \mu - b_j - p_k - b_{jk} - \phi''_k e''_{ijk-1}) = 0$

i.e., $\sum_i y_{ijk} - \frac{y_{...}}{rq} - t\left(\frac{y_{.j.}}{tq} - \frac{y_{...}}{rtq} - \sum_k \phi''_k \frac{e''_{.jk-1}}{tq}\right) -$

$$t\left(\frac{Y_{..k}}{rt} - \frac{y_{...}}{rqt}\right) - t\, b_{jk} - \sum_i \phi''_k e''_{ijk-1} = 0$$

Therefore $\hat{b}_{jk} = \frac{y_{.jk}}{t} - \frac{y_{.j.}}{tq} - \frac{y_{..k}}{rt} + \frac{y_{...}}{rtq} +$

$$\sum_k \phi''_k \frac{e''_{.jk-1}}{t} - \phi''_k \frac{e''_{.jk-1}}{t}$$

$$\frac{\partial R^{(2)}}{\partial \phi''_k} = 0$$

i.e., $\sum_{ij} (y_{ijk} - \mu - b_j - p_k - b_{jk} - \phi''_k e''_{ijk-1})\, e''_{ijk-1} = 0$

$$\sum_{ij} y_{ijk} e''_{ijk-1} - \sum_{ij} \mu e''_{ijk-1} - \sum_j b_j e''_{.jk-1} -$$

$$P_k \sum_{ij} e''_{ijk-1} - \sum_{j} b_{jk}\, e''_{.jk-1} - \sum_{ij} \phi''_k\, e''^2_{ijk-1} = 0$$

$$\sum_{ij} y_{ijk}\, e''_{ijk-1} - \sum_{j} e''_{.jk-1} \left(\frac{y_{.j.}}{tq} - \frac{y_{...}}{rtq} - \sum_{k} \phi''_k \frac{e''_{.jk-1}}{tq} \right) -$$

$$\sum_{j} e''_{.jk-1} \left(\frac{y_{.jk}}{t} - \frac{y_{.j.}}{tq} - \frac{y_{..k}}{rt} + \frac{y_{...}}{rtq} + \sum_{k} \phi''_k \frac{e''_{.jk-1}}{tq} - \right.$$

$$\left. \phi''_k \frac{e''_{.jk-1}}{t} \right) - \phi''_k \sum_{ij} e''^2_{ijk-1} = 0$$

$$\sum_{ij} y_{ijk}\, e''_{ijk-1} - \sum_{j} \frac{y_{.jk}}{t}\, e''_{.jk-1} = \phi''_k \left(\sum_{ij} e''^2_{ijk-1} - \sum_{j} \frac{e''^2_{.jk-1}}{t} \right)$$

Therefore
$$\phi''_k = \frac{\sum_{ij} y_{ijk}\, e''_{ijk-1} - \sum_{j} \frac{y_{.jk}\, e''_{.jk-1}}{t}}{\sum_{ij} e''^2_{ijk-1} - \sum_{j} \frac{e''^2_{.jk-1}}{t}} \quad \text{---(13)}$$

In order to get the ϕ''_k values e''_{ijk-1} values are needed. On substitution of the estimates of parameters except ϕ''_k in the model (12), we get

$$y_{ijk} = \frac{y_{.jk}}{t} - \phi''_k \frac{e''_{.jk-1}}{t} + \phi''_k e''_{ijk-1} + e''_{ijk}$$

Let $z_{ijk} = y_{ijk} - \frac{y_{.jk}}{t}$

Therefore $z_{ijk} = \phi''_k e''_{ijk-1} - \phi''_k \frac{e''_{.jk-1}}{t} + e''_{ijk}$

Since the regression coefficient is zero for the first year, z_{ijk}'s and e''_{ijk}'s are the same for the first year.

For K = 2 to q

$$e''_{ijk} = z_{ijk} - \phi''_k e''_{ijk-1} + \phi''_k \frac{e''_{.jk-1}}{t}$$

Estimates of most of the parameters involve ϕ''_k and that ϕ''_k's themselves involve ϕ''_k's. Therefore ϕ''_k's may be estimated by the method of iteration as was done for model 10. Values of ϕ'_k estimated for the model (8) can be used as the initial solution in this case. Once the estimate of ϕ''_k are obtained iteratively, those of other parameters can be obtained directly.

$$R^{(2)} = \sum_{ijk} (y_{ijk} - \mu - b_j - p_k - b_{jk} - \phi''_k e''_{ijk-1})^2$$

$$= \sum_{ijk} (y_{ijk} - \frac{y_{.jk}}{t} + \phi''_k \frac{e''_{.jk-1}}{t} - \phi''_k e''_{ijk-1})^2$$

$$= \sum_{ijk} y_{ijk} \left(y_{ijk} - \frac{y_{.jk}}{t} + \phi''_k \frac{e''_{.jk-1}}{t} - \phi''_k e''_{ijk-1} \right)$$

$$= \sum_{ijk} y^2_{ijk} - \sum_{ijk} \frac{y_{ijk}\, y_{.jk}}{t} + \sum_{ijk} \phi''_k\, y_{ijk} \frac{e''_{.jk-1}}{t} -$$

$$\sum_{ijk} y_{ijk}\, \phi''_k\, e''_{ijk-1}$$

$$= \sum_{ijk} y^2_{ijk} - \sum_{jk} \frac{y^2_{.jk}}{t} + \sum_{k} \phi''_k \sum_{j} y_{.jk}\, e''_{.jk-1} -$$

$$\sum_{k} \phi''_k \sum_{ij} y_{ijk}\, e''_{ijk-1} \qquad \text{------ (14)}$$

ϕ''_k 's are obtained iteratively, $R^{(2)}$ can be evaluated without much difficulty. Then the treatment sum of squares can be obtained by subracting $R^{(1)}$ from $R^{(2)}$. The various factors are tested against the pooled error mean square derived from the model (2).

The F ratio for testing the treatment differences is therefore given by

$$\frac{(R^{(2)} - R^{(1)})/(t-1)}{R/(q(r-1)\ (t-1)\ -(q-1))}$$

with degrees of freedom (t-1), and (q (r-1) (t-1) - (q-1)) and the F ratio for the treatment x season interaction is

$$\frac{(R^{(1)}-R)/(q-1)\ (t-1)}{R/\ q(r-1)\ (t-1)\ -\ (q-1)}\text{, with degrees of freedom}$$

(q-1) (t-1), q(r-1) (t-1) - (q-1) .

Illustration

ILLUSTRATION

The methodology developed was illustrated using two sets of data. One set of data is generated from an experiment conducted to compare the yield at different levels of pruning in cocoa laid out in RBD in four replications. Observations on yield in four consecutive years, viz. 1989, 1990, 1991, 1992 have been obtained from The Cadbury KAU Co-operative Cocoa Research Project, College of Horticulture, Kerala Agricultural University. The other set of data is obtained by an experiment conducted to compare three varieties of alfalfa laid out in RBD in six replications. Observations on yield in tonnes per acre from cuttings in four consecutive seasons have been taken from Snedecor and Cochran (1967). Both data are given in Appendix-I and II respectively.

A computer programme in Quick Basic was developed to do the analysis described in chapter III and is given in the Appendix

The conventional analysis in the split-plot set up was carried out for two sets of data and are given in Table 1 and 2 respectively.

Table 1. Analysis of variance for data set I

Source	df	SS	MSS	F value
		Split-plot experiment		
Replication	3	195792.46	65264.15	1.79
Factor A (Treatment)	6	268031.86	44671.98	1.2
Error (a)	18	657662.86	36536.83	
Factor B (years)	3	386394.74	128798.25	22.17
Interaction (AB)	18	53961.57	2997.87	0.5161
Error (b)	63	365976.94	5809.16	
Total	111	1927820.42		
		New procedure		
Treatment	6	1383081.00	39718.02	3.63
Treatment x season	18	115723.7	6429.09	
Error	71	111899	10956.32	

Table 2. Analysis of variance for data set II

Source	df	SS	MSS	F value
		Split-plot experiment		
Main plots				
Varieties	2	0.1781	0.0890	0.6534
Blocks	5	4.1499	0.8300	6.0939
Main plots error (a)	10	1.3622	0.1362	
Sub plots				
Date of cutting	3	1.9625	0.6542	23.3642
Intex variety	6	0.2105	0.0351	1.2535
Sub plot error (b)	45	1.2586	0.0280	
		New procedure		
Treatment	2	0.3516	0.1758	3.8052
Treatment x season	6	0.231	0.385	
Error	39	1.8	0.046	

It may be noted that the treatment differences are tested against an error (a) mean square in split-plot analysis having just 18 degrees of freedom in contrast to the error mean square in the new procedure having 71 degrees of freedom. Also the error mean square (a) in the split plot set up is much higher than that in the new methodology. The new procedure is found more sensitive owing to lower mean square error and higher degrees of freedom.

Here also we note that the treatment differences are tested against an error (a) mean square having just 10 degrees of freedom in the split-plot analysis in contrast to the error mean square in the new procedure having 39 degrees of freedom. The error mean square (a) (0.1362) in the split-plot set up is much higher than that in the new methodology (0.0362).

As in the case of data on cocoa, here also the new procedure is found more sensitive owing to lower mean square error and higher degrees of freedom compared to split-plot analysis.

From both the examples, the new procedure is more sensitive than split-plot analysis owing to lower mean square error and higher degrees of freedom.

In case of first example, F-ratio for testing treatment differences in split plot an analysis and in new procedure are 1.2 and 3.63 respectively.

In case of second example, F-ratio for testing treatment differences in split-plot analysis and in a new procedure are 0.6534 and 3.8052 respectively.

In both the examples, it may be noted that F-ratio in the new procedure is significant and F-ratio in the split-plot analysis is non-significant.

Discussion

DISCUSSION

Analysis of data generated from experiments in which observations are taken repeatedly on the same experimental units differs from the usual analysis of data in groups of experiments, so far as the error terms attached to observations on the same experimental unit can't be considered independent. Hence this problem attracted many workers and approximate methodologies have been suggested. All these methods that are in use for this type of analysis attracted criticism. Nevertheless analysis in the split-plot set up is the one widely used due to its low intensity of drawbacks/defects.

Rowell and Walters (1976) criticised the split-plot analysis of experimental data where several successive observations of the same variable have been recorded on each experimental unit on the ground that requirements for such analysis received scant attention and it was often unlikely that these assumptions would be satisfied in experimental situations.

Yates (1982) pointed out that subplot treatments have to be assigned randomly with in each whole plot in split-plot experiments. Since the observations on the same experimental unit at different of time intervals have a temporal sequence, the randomisation for sub-plots cannot be possible. The random assignment of treatments is not carried out have and hence the split-plot set up is incorrect in dealing with repeated measurements.

Gill (1988) concluded that the split-plot analysis for repeated measurements experiments is valid if the error terms attached to the observations on the same experimental units, taken at different points of time have a constant correlation coefficient. But the repeated measurements with a small interval of time age in general highly correlated than those widely separated (Yates, 1982).

The procedure developed here may be viewed in this context. A model which takes the dependence of repeated observations on the same experimental unit into consideration is proposed. The estimation of parameters is tedious in the sense that a closed form solution is not possible. Perhaps, this could have been the reason for suggesting many procedures which are theoretically not very sound. The difficulty in getting straight forward estimation of parameters have been overcome by arriving at numerical solutions by iterative procedures. The tediousness in calculation involved in iterative procedure should not at all be a criteria now because of the availability of computer. In short a procedure of analysis is arrived at which takes the ground reality of the special situation of such experiments into consideration. In this sense the new procedure out shadows all other available competing procedures.

The treatment comparison in split-plot setup is made against error (a) mean square which is in general much higher than the error (b) mean square or to the error variance in a comparable randomised block design. Consequently the precision for comparison of treatments in the split-plot set up is much less compared to the new procedure when the error variance is estimated after eliminating the contribution due to the dependence of error terms. This argument of higher precision is also true with respect to the degrees of freedom by which the error variance is estimated.

It may be noted in the illustrations that the error degrees of freedom in split-plot set up is much lower compared to that in new methodology in both sets of data. Also the error mean square used for testing treatment differences in the split-plot set up is much higher than those in the new methodology in both sets of data. In short, the claim that the new methodology is much more sensitive to distinguish between treatments in comparison with the split-plot analysis is very much evident from the illustrations.

Summary

SUMMARY

The usual method of analysis of data in groups of experiments fails, in situations where observations are taken repeatedly on the same set of experimental units. Such situations arise more often in the case of experiments with perennial crops and animals where the error terms attached to repeated observations on the experimental units can't be assumed independent. Consequently many approximate procedures have been suggested for the purpose, among which, the split-plot approach is the most convincing one and hence the widely used one. But there are many drawbacks for this and hence there are severe criticism against this approach as well.

In order to evolve a procedure that is satisfactory to deal with such situations, a new model which takes the dependence of error terms into account is proposed here. The model can be represented as

$$Y_{ijk} = \mu + t_i + b_j + P_k + t_{ik} + b_{jk} + \phi_k e_{ijk-1} + e_{ijk}$$

where μ is the overall effect

t_i, the effect of the i^{th} treatment

b_j, the effect of the j^{th} replication

P_k, the effect of the k^{th} year

t_{ik}, the interaction effect of the i^{th} treatment and k^{th} year

b_{jk}, the interaction effect of the j^{th} replication and k^{th} year

ϕ_k is the partial regression coefficient of

$(Y_{ijk} - \mu - t_i - b_j - P_k - t_{ik} - b_{jk})$ on e_{ijk-1} and e_{ijk} the error term attached to the observation Y_{ijk} which are assumed independent among themselves and with other terms in the model and normally distributed with mean zero and constant variance say σ_e^2

The parameters of this model are estimated by the method of least squares. For the estimation of treatment sum of squares, a reduced model in which treatment x year interaction is absent was considered for convenience and the principle of least squares applied. In all these cases, application of the principle of least squares did not lead to closed form solutions to the normal equations owing to their complicated nature. Therefore numerical solutions employing on iterative procedure was resorted to obtain the estimates of parameters as well as the various sum of squares required in the analysis of variance.

The method developed is superior to the widely used split-plot approach in two ways. First, the new methodology considers the ground reality of dependence of error terms into considerations and hence free from most of the criticism against the split-plot and others approaches. Second, it is more sensitive in the sense of having more capability of distinguishing between treatments owing to the lower experimental error compared to the split-plot approach.

The new methodology as well as the split-plot approach were illustrated using two different sets of data and the superiority of the former demonstrated.

References

REFERENCES

Aitkin M.I. 1981. Regression models for repeated measurements. Biometrics **37**: 831-832.

Bjornsson H. 1978. Analysis of a series of long term grass land experiments with auto correlated error. Biometrics **34**: 645-651.

Box G.E.P. 1950. Problems in the analysis of growth and wear curger. Biometrics **6**: 362-389.

Cole L.W.L. and **Grizzle J.E.** 1966. Application of multivariate analysis of varience to repeated measurementes experiments. Biometrics **22**: 810 - 823.

Cullis.B.R. and **Ealeson G.K.**, **Wilcox M.E.**, **Laing P.G.** and **Panhuber** 1988. Testing for correlated errors in the analysis of varience. Biometrics **4**: 695-704.

Danford M.B., **Hughes H.M.** and **Mcnee R.C.** 1960. On the analysis of repeated measurements experiments. Biometrics **16**: 547-564

Dempster A.P. 1963. Stepwise multivariate analysis of varience based on principal variables. Biometrics **19**: 478-490.

Diggle P.J. 1988. An approach to the analysis of repeated measurements. Biometrics **44**: 959-974. No.4.

Evans J.C. and **Roberts E.A.** 1979. Analysis of sequential observations with applications to experiments on grazing animals and perennial plants. Biometrics 35: 687 - 693.

Finny D.J. 1956. Multivariate analysis and agricultural experiments. Biometrics **12**: 67-71.

Gaylor D.W. 1978. some properties of the ANOVA with correlated errors with implications for repeated measurements. Biometrics **4**: 727.

Geary P.N. 1989. Modellin the covariance structure of repeated measurements. Biometrics **45**: 1183 - 1195. No.4

Gill J.L. 1986. Repeated measurements. Sensitive test for experiments with four animals. J.anim.Sci. **63**: 943-954.

Gill J.L. 1988. Repeated measurements split plot trend analysis versus analysis of first difference. Biometrics **44**: 289-297

Hearne E.M., **Clark G.M.** and **Hutch J.P.** 1983. Atleast for serial correlation in univariate repeated measures analysis. Biometrics **39**: 237 - 243.

Jenrich R.I. and Schludter **M.D.** 1986. Unbalanced repeated measure models with structural covarience matrices. Biometrics **42**: 505 - 820.

Kenward M. 1984. The use of an autoregressive covarience structure in the analysis of repeated measures. Biometrics **40**: 273.

Kshola R.K., **Rao P.P** and **Das M.N.** 1979. A note on the study of experimental errors in groups of agricultural field experiments conducted in different years. J. Indian Soc. Agric. Stat. **31**: 65 - 68.

Laing D.G. and **Panhuber H.** 1988. Testing for correlated errors in the analysis of varience. Biometrics **44**: 695 - 704. No.3.

Mansour H., **Nordhein E.V.** and Rutledge **J.J.** 1985, maximum likelihood estimation of varience components in repeated measures designs assuring auto regressive errors. Biometrics **41**: 287 - 294.

Pantula S.G. and **Pollock H.H.** 1985. Nested analysis of varience with auto correlatederrors. Biometrics **41**: 909 - 920.

Patterson, 1939. Statical techniques in agricultural research. MacGraw Hill and Co., Newyork.

Raz J. 1989. Analysis of repeated measurements using non parametric smoother's and randomization tests. Biometrics **45**: 851-871

Rowell J.G. and **Walter D.E.** 1976. Analysing data with repeated observations on experiment unit. J. Agric. Sci. **87:423-432**

Smith H., **Ganadesikan R.** and **Hugles J.B.** 1962. Multivariate analysis of variance. Biometrics **18**: 22 - 41.

Snedecor G.W. and **Cochran W.G.** 1967. statistical methods. 6th Editions Oxford and IBH Publishing Co.

Steel, R.P.G. 1955. An analysis of perennial crop data. Biometrics **11**: 201 - 212.

Steel, R.D.G. and **Torrie.J.H.** 1960. Principles and procedures of analysis. McGraw Hill & Co., Newyork.

Sukumaran K. 1991. Pooled analysis of dependent sets of data. M.Sc thesis, Kerala Agricultural University, Vellanikkara.

Wallenstein S. 1982. Regression models for repeated measurements. Biometrics **38**: 849 - 850.

Yates F. and **Cochran W.G.** 1938. The analysis of groups of experiments. J. Agric. Sci. **28**: 556 - 580.

Yates F. 1954. The analysis of experiments containing different crop rotations. Biometrics **10**: 329 - 348.

Yates F. 1982. Regression models for repeated measurements. Biometrics **38**: 850 - 853.

Appendices

Appendix I

Yield data pruning trial in cocoa

Years		T_1	T_2	T_3	T_4	T_5	T_6	T_7
I	R_1	327	583	241	306	642	312	416
	R_2	198	146	247	322	333	218	477
	R_3	44	190	253	461	228	268	551
	R_4	297	45	270	158	383	135	127
II	R_1	225	409	270	276	201	303	349
	R_2	175	187	94	380	382	294	187
	R_3	139	54	314	457	402	306	547
	R_4	212	183	169	103	312	105	169
III	R_1	90	258	94	214	134	162	202
	R_2	72	96	89	181	134	68	153
	R_3	85	26	134	296	239	192	244
	R_4	216	90	69	86	213	36	82
IV	R_1	168	440	162	287	344	193	456
	R_2	121	162	148	305	215	199	268
	R_3	106	60	240	287	235	289	485
	R_4	430	220	109	70	292	79	158

Appendix II

Date	Variety/ replication	1	2	3	4	5	6	
	Ladack	2.17	1.88	1.62	2.34	1.58	1.66	11.25
A	Cossack	2.33	2.01	1.70	1.78	1.42	1.35	10.59
	Ranger	1.75	1.95	2.13	1.78	1.31	1.30	10.22
		6.25	5.84	5.45	5.90	4.31	4.31	32.06
	Ladack	1.58	1.26	1.22	1.59	1.25	0.94	7.84
B	Cossack	1.38	1.30	1.85	1.09	1.13	1.06	7.81
	Ranger	1.52	1.47	1.80	1.37	1.01	1.31	8.48
		4.48	4.03	4.87	4.05	3.39	3.31	24.13
	Ladack	2.29	1.60	1.67	1.91	1.39	1.12	9.98
C	Cossack	1.86	1.70	1.71	1.54	1.67	0.88	9.46
	Ranger	1.5	1.61	1.82	1.56	1.23	1.13	8.90
		5.70	4.91	5.30	5.01	4.29	3.13	28.34
	Ladack	2.23	2.01	1.82	2.10	1.66	1.10	10.92
D	Cossack	2.27	1.81	2.01	1.40	1.31	1.06	9.86
	Ranger	1.56	1.72	1.99	1.55	1.51	1.38	9.66
		6.06	5.54	5.82	5.05	4.48	3.49	30.44

ITERATIVE PROCEDURE

```
10  REM TO FIND THE VALUE OF AN EQUATION
20  DIM X(10,10,10), Z(10,10,10), E(10,10,10), AB(10,10),
    AC(10,10), BC(10,10), EEI(10)
30  DIM EB(10,10), E(10,10), A(10), B(10), C(10), EEJ(10,10),
    EEI(10,10), B1(10), B2(10)
40  DIM SSI (10)
50  REM INPUT "NUMBER OF TREATMENTS (1)", NT
60  REM INPUT "NUMBER OF REPLICATIONS (J)", NR
70  REM INPUT "NUMBER OF YEARS (K)", NY
80  REM INPUT "INPUT DATA FILE NAME", N$
81  NT=3; NR=6; NY=4; N$="PR"
82  OPEN"0", #2, "POUT"
90  OPEN "i",#1,N$
100 Y1=0
110 FOR K=1 TO NY
120 C(K)=0; NEXT K
130 FOR K=1 TO NY
140 FOR I =1 TO NT
150 AC(I,K)=0; NEXT I:NEXT K
160 FOR J=1 TO NR
170 FOR K=1 TO NY
180 BC(J,K)=0: EA(J,K)=0
190 NEXT K:NEXT J
200 FOR I=1 TO NT
210 A(I)=0; NEXT I
220 FOR I=1 TO NT
230 FOR J=1 TO NR
240 AB(I,J)=0; NEXT J:NEXT I
250 FOR K=1 TO NY:FOR I=1 TO NT:FOR J=1 TO NR
260 INPUT #1,X(I,J,K)
270 Y1=Y1+X(I,J,K):AB(I,J)=AB(I,J)+X(I,J,K)
```

```
280  BC(J,K)=BC (J,K) + X (I,J,K):AC(I,K)=AC(I,K)+X(I,J,K)
290  A(I)=A(I) + X(I,J,K):C(K)=C(K)+X(I,J,K)
300  NEXT J:NEXT I:NEXT K
310  FOR K=1 TO NY:FOR I=1 TO NT:FOR J=1 TO NR
320  Z(I,J,K)=X(I,J,K)-C(I,K)/NR-BC(J,K)/NT+(C(K)/(NI*NR)
330  NEXT J:NEXT I:NEXT K
340  RS1=0
350  FOR K=2 TO NY
360  S1=0:SS1=0:S12=0
370  FOR I=1 TO NT : FOR J=1 TO NR
380  SS1=SS1+Z(I,J,K-1)*Z(I,J,K-1)
390  S12=S12+X(I,J,K)*Z(I,J,K-1)
400  NEXT J:NEXT I
410  B(K)=S12/SS1
420  FOR I=1 TO NT
430  FOR J=1 TO NR
440  Z(I,J,K)=Z(I,J,K)-B(K)* Z(I,J,K-1)
450  NEXT J:NEXT I
460  PRINT #2,"PHI(":K,")=",B(K)
470  NEXT K
480  FOR K=1 TO NY:FOR I=1 TO NT:FOR J=1 TO NR
490  EA(J,k)=EA(J,K)+Z(I,J,K)
500  RS1=RS1+Z(I,J,K)*Z(I,J,K)
510  NEXT J:NEXT I:NEXT K
520  PRINT #2,"RESIDUAL SS=",RS1
530  REM ESTIMATAION OF TREATMENT MEANS
540  FOR I=1 TO NT
550  TA=A(I)/(NR*NY) -Y1/(NR*NY*NT)
560  PRINT #2, "T(",I,")=",TA
570  NEXT I
580  FOR K = 1 TO NY
590  FOR I = 1 TO NT
600  FOR J = 1 TO NR
```

```
610  Z(I,J,K) = X(I,J,K) -(A(I)/(NY*NR))
     -(BC(J,K)/NT)+(Y1/(NY*NR*NT))
620  EEJ (I,K) = EEJ(I,K)+Z(I,J,K)
630  EEI (J,K)=EEI(J,K)+Z(I,J,K)
640  IF K>1 THEN 660
650  E(I,J,K) = Z(I,J,K)
660  NEXT J:
670  NEXT I:NEXT K
680  KV=0
690  FOR I=1 TO NT
700  EE1(I)=0
710  FOR K=1 TO NY
720  EB(I,K)=0 : SSI(K)=0
730  IF K=1 THEN EB(I,K)=EEJ(I,K)
740  NEXT K:NEXT I
750  FOR I=1 TO NT
760  FOR K=2 TO NY
770  EE1 (I) = EE1 (I) +B(K)*EEJ(I,K-1)
780  NEXT K:NEXT I
790  S1 = 0
800  FOR J=1 TO NR
810  FOR K=1 TO NY
820  IF K>1 THEN 840
830  EA(J,K)=EEI(J,K)
840  EA(J,K)=0:NEXT K:NEXT J
850  FOR K = 1 TO NY
860  FOR I = 1 TO NT
870  FOR J = 1 TO NR
880  E(I,J,K) = Z(I,J,K)-B(K)*E(I,J,K-1)+EE1(I)/(NR*NY)
890  EB(I,K) = EB(I,K) + E(I,J,K)
900  EA(J,K) = EA(J,K) + E(I,J,K)
910  NEXT J
911  S11=0:FOR KM=2 TO NY
920  S11=S11+EB(I,KM-1)*B(KM) :NEXT KM
```

```
921  S1=S1+S11*EB(I,K-1)
930  NEXT I:NEXT K:
940  FOR K=1 TO NY
950  FOR J=1 TO NR
960  SSI(K) = SSI(K)+EA(J,K)*EA(J,K)
970  NEXT J: NEXT K:
980  FOR K=2 TO NY
990  PH1 = 0: PH2 = 0: SSE=0
1000 FOR I = 1 TO NT
1001 PH2=PH2+ (A(I)*EB(I,K-1))/(NR*NY)
1010 FOR J = 1 TO NR
1020 PH1 = PH1+X(I,J,K)*E(I,J,K-1)
1050 SSE=SSE+E(I,J,K-1)*E(I,J,K-1)
1060 NEXT J:NEXT I
1070 B1(K) = (PH1 - PH2+S1/(NR*NY))/SSE
1080 PRINT #2, B1(K):NEXT K
1090 KV=KV+1
1100 PRINT #2,"CALCULATION FOR ITERATION", KV, "OVER"
1110 FOR K = 2 TO NY
1120 IF ABS(B1(K)-B(K)) >.001 THEN 1150
1130 NEXT K
1140 GOTO 1210
1150 FOR K=2 TO NY
1160 B(K) = B1 (K)
1170 NEXT K
1180 FOR I=1 TO NT:FOR K=1 TO NY
1190 EEJ(I,K)=EB(I,K):NEXT K:NEXT I
1200 GOTO 690
1210 SSZ=0
1220 FOR K=1 TO NY:FOR I=1 TO NT:FOR J=1 TO NR
1230 SSZ=SSZ+E(I,J,K)*E(I,J,K):EA(J,K)=0 :SSI(K)=0
1240 NEXT J:NEXT I:NEXT K
1250 REM CALCULATION FOR REDUCED MODEL OVER
1260 PRINT #2. "SSZ=",SSZ
```

```
1270  SSE1=0:FOR K=1 TO NY:SS1(K)=0:FOR J=1 TO NR:
      EEI(J,K)=0:NEXT J:NEXT K
1280  FOR K=1 TO NY:FOR I=1 TO NT:FOR J=1 TO NR
1290  Z(I,J,K)=X(I,J,K)+(BC(J,K)/NT)
1300  EEI(J,K)+EEI(J,K)+Z(I,J,K)
1310  IF K>1 THEN 1360
1320  E(I,J,K)=Z(I,J,K)
1330  SSE1=SSE1+E(I,J,K)*E(I,J,K)
1340  SSI(K)=SSI(K)+E(I,J,K)*E(I,J,K)
1360  NEXT J
1370  NEXT I:NEXT K
1380  KV=0
1390  FOR J=1 TO NR
1400  EE1(J)=0
1410  FOR K=1 TO NY:EA(J,K)=0
1420  IF K=1 THEN EA(J,K)=EEI(J,K)
1430  NEXT K:NEXT J
1440  SSE=SSE1
1450  FOR J=1 TO NR:FOR K=2 TO NY
1460  EE1(J)=EE1(J)+B1(K)*EEI(J,K-1)
1470  NEXT K:NEXT J
1480  S1=0
1490  FOR K=2 TO NY:FOR J=1 TO NR:FOR I=1 TO NT
1500  E(I,J,K)=Z(I,J,K)-B1(K)*E(I,J,K-1)
1510  EA(J,K)=EA (J,K) + E(I,J,K)
1520  SSE=SSE+E(I,J,K)*E(I,J,K)
1530  SS1(K)=SSI(K)+E(I,J,K)*E(I,J,K)
1540  NEXT I
1550  S1=S1+EA(J,K-1)*EA(J,K-1)*B1(K)
1560  NEXT J:NEXT K
1570  FOR K=2 TO NY
1580  PH1=0:PH2=0
1590  FOR J=1 TO NR:FOR I=1 TO NT
1600  PH1=PH1+X(I,J,K)*E(I,J,K-1):NEXT I
```

```
1610  NEXT J
1620  B2(K)=(PH1-PH2+(S1/NT))/SSI(K-1)
1630  PRINT #2(K):NEXT K
1640  KV=KV+1
1650  PRINT "CALCULATION FOR ITERATION",KV,"OVER"
1660  FOR K=2 TO NY
1670  IF ABS (B2(K)=B1(K))>.001 THEN 1700
1680  NEXT K
1690  GOTO 1770
1700  FOR K=2 TO NY
1710  B1(K)=B2(K)
1720  NEXT K
1730  FOR J=1 TO NR :FOR K=1 TO NY
1740  EEI(J,K)=EA(J,K):NEXT K:NEXT J
1750  GOTO 1390
1770  PRINT #2,"SSE =",SSE
1780  PRINT #2."TREATMENT SS=",SSE-SSZ
1781  CLOSE
1790  END
```

ABSTRACT

Analysis of variance model for the groups of experiments needs modification, when observations are taken repeatedly on the same experimental units owing to the autocorrelated nature of error terms. A model which takes the dependence of error terms into consideration was evolved for dealing such situations. But estimation of parameters using least square principle and their tests of significance not straight forward. Therefore numerical solutions using iterative technique was employed for estimation of parameters of the model.

The newly developed procedure was compared to the widely used analysis of the split-plot setup and the comparative advantage of the new method was established.

The new methodology along with the widely used analysis of the split-plot set up were illustrated using two different sets of data. The superiority of the new method over the split-plot analysis was demonstrated in both sets of data.