POOLED ANALYSIS OF DEPENDENT SETS OF DATA

By

SUKUMARAN, K.

THESIS

Submitted in partial fulfilment of the requirement for the degree

Master of Science in Agricultural Statistics

Faculty of Agriculture
Kerala Agricultural University

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

1991

To my loving parents

DECLARATION

I hereby declare that this thesis entitled "Pooled analysis of dependent sets of data" 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,
27-7-1991.

SUKUMARAN, K.

CERTIFICATE

Certified that this thesis entitled "Pooled analysis of dependent sets of data" is a record of research work done independently by Sri. SUKUMARAN, K. under my guidance and supervision and that it has not previously formed the basis for the award of any degree, fellowship or associateship to him.

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

Vellanikkara,
27.7.1991

CERTIFICATE

We, the undersigned, members of the Advisory Committee of Sri. **SUKUMARAN, K.**, a candidate for the degree of Master of Science in Agricultural Statistics, agree that the thesis entitled "Pooled analysis of dependent sets of data" may be submitted by Sri. **SUKUMARAN, K.**, in partial fulfilment of the requirement for the degree.

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

Dr. K.C. George, 13-8-91
Professor & Head
Department of Statistics,
College of Veterinary and
Animal Sciences, Mannuthy.
(Member)

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

Dr. P.A. Wahid
Professor (Radiotracer)
Radiotracer Laboratory
Kerala Agricultural University
Vellanikkara
(Member)

EXTERNAL EXAMINER 13/8/91

ACKNOWLEDGEMENTS

I express my deep sense of gratitude and indebtedness to Sri. V.K. Gopinathan Unnithan, Associate Professor of Agricultural Statistics 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 are due to Dr. K.C.George, Professor and Head, Department of Statistics, College of Veterinary and Animal Sciences, Mannuthy; Sri. P.V. Prabhakaran, Professor and Head, Department of Agricultural Statistics, College of Horticulture, Vellanikkara; Dr. P.A. Wahid, Professor (Radio-tracer), Radiotracer Laboratory, Kerala Agricultural University, Vellanikkara; members of the Advisory Committee for the painstaking involvement throughout the course of the programme.

I also take this occasion to express my deep sense of gratefulness to Sri. Abdul Rasak, Associate Professor, Department of Statistics, College of Veterinary and Animal Sciences, for his critical suggestions during the course of this investigation.

I would like to express my sincere thanks to Dean, College of Veterinary and Animal Sciences and the Associate Dean, College of Horticulture 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 much obliged to Babu Ratnam, Seena, C, Chandrika, C. for their valuable help and co-operation in completing my research work successfully.

I am grateful to the Kerala Agricultural University for the fellowship awarded to me during the course of research work. I am highly obliged to all my friends and wellwishers for their assistance and kind co-operation.

I must also register wholehearted appreciation to Sri.V.T.Kurian who typed out the entire work with sincerity and accuracy.

Lastly, but not least I express my indebtedness to my parents, brothers and sister who are always being a source of inspiration and encouragement to me in all my endeavours.

(SUKUMARAN, K.)

CONTENTS

Introduction

INTRODUCTION

In large scale experimental programmes it is necessary to repeat a trial of a set of treatments at a number of places and/or at a number of years. The aim of such repetition is to study the susceptibility of the treatment effects to places and climatic variations. The results of a single experiment conducted in any particular year cannot be relied upon as they are subjected to the climatic conditions which fluctuate from year to year at any place. Hence to draw reliable conclusions relevant for at least a few years to come, it is necessary to repeat the experiment in a number of years.

In applied sciences, observations are often collected on some aspects of an individual under different experimental conditions over time. Experiments in which trees, animals or human subjects are assigned randomly to treatments and then measured repeatedly at selected intervals of time are common in many scientific disciplines.

For drawing valid conclusions in experiments with repeated measurements, it becomes necessary to make a joint statistical analysis of the data combining observations of the individual years or seasons. Proper randomisation of treatments to experimental units ensures independence of error terms in the analysis of variance model. But in experiments with repeated measurements, randomisation remains unchanged year after year and the observations in successive years or seasons in the same individual or experimental unit can no longer be considered independent. In an experiment with repeated measurements, the measurements have a temporal sequence with the consequence that measurements on the same subject separated in a small time interval will in general be highly correlated. This introduces correlation among the error terms. These type of auto-correlation among the error terms of successive years will definitely affect the precision of overall treatment comparison. Since the error terms from year after year are dependent, the assumption needed for the analysis of variance are not satisfied and hence the analysis of variance cannot be applied directly for such type of data.

A solution suggested by many workers in dealing with such experiments is to consider the experiment as a split-plot arrangement with years or seasons on subplots, within each treatment main plot. But split-plot design requires the random arrangement of sets of subplot treatments within each main plot and that cannot be expected in the case of trials repeated over several seasons. In such experiments we have to confront with a systematic arrangement of seasons in chronological order under each main plot. Here also the assumption of independence of error terms does not seem to be wholly valid. Even if we assume that the sets of experimental years constitute a random sample of years for a population of years, the systematic occurance of seasons makes the estimate biased.

At present, there is no satisfactory procedure to have an overall analysis of data taken repeatedly from the same set of experimental units. This is a major handicap in the analysis of data generated from experiments on perennial crops and animals. A satisfactory solution to this problem will help in getting many inferences from such experiments.

In view of the circumstances the present investigation is therefore formulated to evolve a procedure to analyse data generated by repeated observations on the same set of experimental units.

Review of Literature

2. REVIEW OF LITERATURE

The comprehensive studies on the various problems of data analysis in experiments involving repeated measurements are scanty in literature. In an experiment 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. A review of the available literature on the subject is furnished below under the sub headings General, Split-plot set up, Analysis of differences, Multivariate analysis, ARMA models and Non parametric methods.

General

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 standard analysis of variance procedure suitable for dealing with the results of the single experiment needed modification, owing to the lack of equality in the error components and that in the interactions of different groups of treatments with places or time or both.

Yates (1954) investigated the problem arising in the analysis of data from long term experiments containing different crop rotations. The method was illustrated by application to rice-pasture experiment containing rotation of different lengths and with different proportions of rice to pasture. When the design of the experiment was such that each block contained plots which some times carried a given crop but did not all carry the crop in the same set of years, the year and block totals were not found to be orthogonal to the plot totals. He suggested the method of fitting contrasts to obtain separate estimates of plot error and plot x year error which were free from year x block interactions.

Khosla et al. (1979) studied the behaviour of experimental errors and presence of treatments x year interaction in the case of groups of experiments, involving single experimental error, conducted at different research stations in the state of Gujarat during different years of the period 1960-65. Homogeneity or otherwise of experimental errors was studied with reference to different crops, types of experiments and broad soil types. Results of 199 groups of experiments conducted at different agricultural research stations in the state of Gujarat on different crops were considered for studying the behaviour of experimental errors.

They used the weighted analysis for testing the presence of treatment x year interaction which made the testing of overall treatment effects and observed that the interaction of treatment x year was presented in 35.7 per cent of the cases irrespective of the homogeneity or heterogeneity of error variance.

Cullis and McGilchrist (1990) developed the model

$$y_{ijk} = \mu_k + \theta_{jk} + G_{ijk} + M_{ijk}$$

for growth data from designed experiment where

μ_k represents the average relative growth from t_{k-1} to t_k. θ_{jk} represents the additional effect of treatment j from t_{k-1} to t_k. G_{ijk} is the 'growth error' of individual i, j from time t_{k-1} to t_k, ie. departure of this individual from the average growth for the treatment. Such errors were found to be correlated over time but individuals were assumed to be independent. M_{ijk} is the first differenced measurement or sampling errors which were taken to be independent but the differencing operation imparts a correlation to the derived terms. Residual maximum likelyhood (REML) was used to estimate the parameters of the model. The model was also extended to incomplete data and was shown to over come some of the practical difficulties encountered with the profile model. The procedure was applied to data from experiments on pigs and sheep.

Split-plot set up

Patterson (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.

In some experiments successive observations are made in the same unit over a period of time. Steel and Torrie (1960) observed that such data are analogous to those from split-plot design in many aspects and their analysis is often conducted as such.

Pearce (1953) proposed split-plot design for the analysis of data from perennial crop experiments with periods assigned to the sub-plots and treatments assigned to the whole plots.

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 likelyhood procedure in presence of serial correlation.

Rowell and Walter (1976) found that the assumptions required for the analysis of data from a group of experiments involving same treatments carried out a number of years by using split plot analysis would not be satisfied in experimental situations. They suggested an alternative method in which contrast over time are analysed and found that such analysis are always valid, computationally and readily interpretable. This may also be used to guage the validity of the splits plot analysis.

Gill (1986) proposed some modification in split-plot analysis for the repeated measurements when the number of animals per treatment is not more than five or six. He partitioned the treatment x period interaction of the univariate split-plot analysis to permit sensitive comparison of treatments. Modifications for the procedure were given for the case of heterogeneous variance and covariance.

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.

Yates (1982) reported that the split-plot analysis suggested by Aitkin for an experiment in which rectal temperature were measured on a group of 10 subjects at 20 minute intervals over exposure period of two hours, at four different ambient temperatures was an incorrect terminology. He observed that in experiments with repeated measurements the measurements have a temporal sequence, with the consequence that measurements on the same subject separated by a small time interval will in general be more highly correlated than those more widely separated. He found that the six values for any subject exposed to a particular ambient temperature could be fitted exactly by a fifth degree polynomial.

Analysis of differences

Box (1950) observed that difference of successive observations atleast resulted in a very simple covariance pattern for the errors and then the analysis by a simple application of the techniques of the analysis of variance. He suggested analysis of increments of response relevant for repeated measurements. A multivariate extension of the analysis of variance suggested for situations where simple set up was inappropriate.

Gill (1988) criticised the conclusion by Box (1950) that analysis of increments of response was superior to trend analysis of the original data with split-plot model as spurious. He was of the opinion that reduction of interperiod correlation by using first differences did not necessarily eliminate problems with heterogeneity of the variance covariance matrix over time. For the homogeneous conditions, the expected variance of a simple trend contrast (between two treatments, for adjoined periods) was shown to be the same for either analysis, but the analysis of increments incurred a loss of degrees of freedom which could be critical in studies with few experimental units.

Multivariate approach

Steel (1955) suggested multivariate analysis of yield with observations for a plot in different years forming the vector variable for forage crops where varieties were grown in the same plots in all years. The tests of significance used are quite valid even when the error terms in succeeding years are correlated and or even when residual variations are heterogeneous with respect to year. He applied a bivariate analysis of variance to perennial crop data and made a test of hypothesis about the varietal effect and compared the result to that of univariate analysis.

Finney (1956) severely criticised the use of multivariate analysis of variance, construction of cannonical variates for the analysis and the interpretation of agricultural experiments repeated over years. He observed that the method described by Steel (1955) to be tedious. He 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.

Danford et al (1960) studied an experiment involving repeated measurements on the same individuals over time. They tested the assumptions and techniques of the usual univariate analysis of variance procedure and observed that the assumptions of equal variance and covariance were not fulfilled by the data. He suggested the multivariate procedures where valid univariate analysis is not justified. It was shown that the univariate and multivariate tests were identical assymptotically.

Dempster (1963) extended the step wise testing methods of multivariate analysis of variance to the linear combinations of variables resulting from principal component analysis.

Cole and Grizzle (1966) used multivariate analysis of variance to repeated measurements experiments. They were of the opinion that the multivariate analysis provided a unified approach to the analysis of data from repeated measurements with all the power and flexibility of the univariate analysis of variance.

Morrison (1972) performed the test of the repeated measurements hypothesis of equality of the elements of multivariate normal mean vector under the assumptions of positive definite, symmetric, reducible, reducible to the compound symmetry pattern variance covariance matrix. The test included an alternative form of the usual Hotellings Hsu Statistic. The efficiencies of the special methods were measured in terms of average squared length of their simultaneous confidence intervals.

ARMA models

The errors in a group of experiments were identified as second order auto regressive AR(2), by Bjornsson (1978) and error structure was adequately represented by auto regressive moving average model (ARMA (p, q). He found that the residuals in perennial crop experiments were auto correlated and the auto correlation normally decreased as the time interval increased.

Yang and Carter (1983) studied the problem of testing the null hypothesis of equality of the group means when the successive observations over time were made on individual subjects classified into different groups. It was assumed that ther was a random effect for each individual and that successive observations on each individual followed a simple ARMA model. They pointed out that in many practical situations usual analysis of variance F test, performed on the average as an efficient test.

Non parametric method

The method described by Rai and Rao (1978) used only information on 'order' and made no use of the actual values of the variate. For this reason no assumption is required to be made as to the nature of underlying universe. The method is thus applicable to a wide class of problems to which the analysis of variance cannot validly be applied. The theoretical discussion of the efficiency of this procedure relative to the analysis of variance indicated that in situations where the latter method could validly be applied and when the number of sets of ranks was large, the maximum loss of information in this method was only 36 per cent.

Methodology

METHODOLOGY

Pooled analysis of data generated from multilocational trials have been developed and is in wide use satisfactorily. But the available procedure cannot be used directly when the error terms in the analysis of variance model cannot be assumed independent a situation characteristic of experiments generating information on repeated measurements. Since the error terms in the conventional analysis of variance model are not independent in such situations, we shall 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 the same experimental units for q years. The model can easily be adopted to other types of designs with appropriate modifications.

Let the observation from the experimental unit in the j^{th} replication receiving i^{th} treatment at the k^{th} year, say y_{ijk} be represented by

$$y_{ijk} = \mu + t_i + b_j + p_k + t_{ik} + b_{jk} + \sum_{n=2}^{k} \beta^{k}_{n-1} e_{ijn-1} + e_{ijk} \quad \ldots \quad (3.1)$$

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}, 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} is 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 $\sigma_e{}^2$

We shall estimate the parameters in the model by the method of least squares. The error sum of square is given by

$$R = \sum_{ijk}(y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \sum_{n=2}^{k} \beta^k_{n-1} e_{ijn-1})^2 \qquad \ldots \quad (3.2)$$

The summation being over all possible values of i, j and k. Estimates of the parameters μ, t_i, b_j, p_k, t_{ik}, b_{jk} and β^k_p 's can be obtained by solving the set of normal equations derived by differentiating R with respect to each of these parameters and equating to zero. Thus differentiating R with respect to μ , and equating to zero, we have

$$\frac{\partial R}{\partial \mu} = 0$$

ie. $- 2 \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1}) = 0$

... (3.3)

Since the number of independent normal equations obtained will be less than the number of parameters to be estimated, we impose the conditions

$$\sum_{i=1}^{t} t_i = \sum_{j=1}^{r} b_j = \sum_{k=1}^{q} p_k = \sum_{i} t_{ik} = \sum_{k} t_{ik} = \sum_{j} b_{jk} = \sum_{k} b_{jk} = 0$$

... (3.4)

As a consequence $\sum_{ij} e_{ijl} = 0, \; l = 1, 2, \ldots q$

Therefore the estimator of μ, from(3.3) is given by

$$\hat{\mu} = \frac{y_{\ldots}}{qrt} \qquad \ldots \quad (3.5)$$

where, $y_{\ldots}$ is the sum of y_{ijk} over all possible values of i, j and k

Differentiating R with respect to t_i and equating to zero, we get

$$\frac{\partial R}{\partial t_i} = 0$$

ie. $-2 \sum_{jk} (y_{ijk} - \mu - t_i - b_j - p_k - t_{ik} - b_{jk} - \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1}) = 0$

$$\sum_{jk} y_{ijk} - qr\mu - qrt_i = 0$$

Since it can be shown that $\sum_{i} e_{ijl} = \sum_{j} e_{ijl} = 0$

$l = 0, \ldots q$

Therefore,

$$\hat{t}_i = \frac{y_{i..}}{qr} - \frac{y_{...}}{qrt} \qquad \ldots \qquad (3.6)$$

where $y_{i..}$ is the sum of y_{ijk} overall possible values of j and k.

estimation of b_j

Differentiating R with respect to b_j and equating to zero, we get

$$\frac{\partial R}{\partial b_j} = 0$$

ie. $\sum_{ik} y_{ijk} - qt\mu - qtb_j = 0$

$$\hat{b}_j = \frac{y_{.j.}}{qt} - \frac{y_{...}}{qrt} \qquad \ldots \qquad (3.7)$$

where $y_{.j.}$ is the sum of y_{ijk} overall possible values of i and k.

Estimation of P_k

Differntiating R with respect to p_k and equating to zero, we get

$$\frac{\partial R}{\partial P_k} = 0$$

ie. $\sum_{ij} y_{ijk} - rt\mu - rtP_k = 0$

$$\hat{P}_k = \frac{y_{..k}}{rt} - \frac{y_{...}}{qrt} \qquad \ldots \qquad (3.8)$$

where $y_{..k}$ is the sum of y_{ijk} over all possible values of i and j

Estimation of t_{ik}

Differentiating R with respect to t_{ik} and equate to zero, we have

$$\frac{\partial R}{\partial t_{ik}} = 0$$

ie. $\sum_j y_{ijk} - r\mu - rt_i - rP_k - rt_{ik} = 0$

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

$$= \frac{y_{i.k}}{r} - \frac{y_{...}}{qrt} - \frac{y_{i..}}{qr} + \frac{y_{...}}{qrt} - \frac{y_{..k}}{rt} + \frac{y_{...}}{qrt}$$

$$= \frac{y_{i.k}}{r} - \frac{y_{i..}}{qr} - \frac{y_{..k}}{rt} + \frac{y_{...}}{qrt} \quad \ldots \quad (3.9)$$

where $y_{i.k}$ is the sum of y_{ijk} overall possible values of j

Estimation of b_{jk}

Differentiating R with respect to b_{jk} and equating to zero, we have

$$\frac{\partial R}{\partial b_{jk}} = 0$$

ie. $\sum_i y_{ijk} - t\hat{\mu} - t\hat{b}_j - t\hat{P}_k - t\hat{b}_{jk} = 0$

$$\hat{b}_{jk} = \frac{y_{.jk}}{t} - \hat{\mu} - \hat{b}_j - \hat{P}_k \quad \ldots \quad (3.10)$$

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

$$\hat{b}_{jk} = \frac{y_{.jk}}{t} - \frac{y_{.j.}}{qt} - \frac{y_{..k}}{rt} + \frac{y_{...}}{qrt} \quad \ldots \quad (3.11)$$

where $y_{.jk}$ is the sum of y_{ijk} over all possible values of i.

<u>Estimation of β_p^k</u>

Differentiating R with respect to β_p^k and equating to zero, we have

$$\frac{\partial R}{\partial \beta_p^k} = 0$$

ie. $$\sum_{ij} y_{ijk} e_{ijp} - \sum_{ij} \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1} e_{ijp} = 0 \quad \ldots (3.12)$$

$p = 1,2 \ldots k-1$

$k = 2,3 \ldots q$

$$\sum_{ij} y_{ijk} e_{ijp} = \sum_{ij} \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1} e_{ijp}$$

since e_{ijk}'s are assumed independent among themselves and with other terms in the model, $\sum_{ij} e_{ijl} e_{ijm} = 0$; $m \neq l = 1,2 \ldots q$.

So the estimate of β_p^k is

$$\beta_p^k = \frac{\sum_{ij} y_{ijk} e_{ijp}}{\sum_{ij} e_{ijp}^2}, \quad p = 1,2 \ldots k-1; \quad k = 2,3 \ldots q.$$

In order to get the estimates of β_p^k, e_{ijp} are to be substituted by their estimates.

On substitution of the estimates of the parameters except β_p^k's in the model (3.1) we get

$$y_{ijk} = \frac{y_{i.k}}{r} + \frac{y_{.jk}}{t} - \frac{y_{..k}}{rt} + \sum_{l=1}^{k-1} \hat{\beta}_l^k e_{ijl} + e_{ijk} \quad \ldots \quad (3.14)$$

$i = 1,2 \ldots t$

$j = 1,2 \ldots r$

$k = 1,2 \ldots n-1$

$n = 2,3 \ldots q$

let

$$z_{ijs} = y_{ijs} - \frac{y_{i.s}}{r} - \frac{y_{.js}}{t} + \frac{y_{..s}}{rt} \quad \ldots \quad (3.15)$$

$i = 1,2 \ldots t$

$j = 1,2 \ldots r$

$s = 1,2 \ldots n-1$

$n = 2,3 \ldots q$

Using the notations of (3.15), the equation (3.14) can be rearranged to get

$$\begin{bmatrix} z_{ij1} \\ z_{ij2} \\ \vdots \\ z_{ijn-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ \beta_1^2 & 1 & 0 & 0 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \beta_1^{n-1} & \beta_2^{n-1} & \beta_3^{n-1} & \beta_4^{n-1} & \ldots & \beta_{n-2}^{n-1} & 1 \end{bmatrix} \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \\ e_{ijn-1} \end{bmatrix} \quad \ldots (3.16)$$

ie. $I = BE$

where

$$I = \begin{bmatrix} z_{ij1} \\ z_{ij2} \\ \vdots \\ z_{ijn-1} \end{bmatrix} ; \quad B = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \beta_1^2 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \beta_1^{n-1} & \beta_2^{n-1} & & \cdots & \beta_{n-2}^{n-1} & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijn-1} \end{bmatrix}$$

Denoting the inverse of B as A

ie.

$$\begin{bmatrix} \beta_1^1 & 0 & \cdots & 0 \\ \beta_1^2 & \beta_2^2 & \cdots & 0 \\ \beta_1^{n-1} & \beta_2^{n-1} & \cdots & \beta_{n-1}^{n-1} \end{bmatrix}^{-1} = \begin{bmatrix} a_1^1 & 0 & \cdots & 0 \\ a_1^2 & a_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_{n-1}^{n-1} \end{bmatrix} \quad (3.17)$$

Where $a_j^i = \sum_{k=j}^{i-1} \beta_k^i \, a_j^k$, $j = 1,2 \ldots i-1; \; i = 2,3 \ldots n-1$

and $a_i^i = 1$, $i = 1, 2, \ldots . n-1$

Therefore

$$\begin{bmatrix} e_{ij1} \\ e_{ij2} \\ \vdots \\ e_{ijn-1} \end{bmatrix} = \begin{bmatrix} a_1^1 & 0 & 0 & \cdots & 0 \\ a_1^2 & a_2^2 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_1^{n-1} & a_2^{n-1} & \cdots & & a_{n-1}^{n-1} \end{bmatrix} \begin{bmatrix} z_{ij1} \\ z_{ij2} \\ \vdots \\ z_{ijn-1} \end{bmatrix} \quad ..(3.18)$$

So that

$$\hat{e}_{ijk} = a_1^k z_{ij1} + a_2^k z_{ij2} + \cdots + a_k^k z_{ijk}$$

$$= \sum_{s=1}^{k} a_s^k z_{ijs} \quad \cdots \quad (3.19)$$

Substituting (3.19) in (3.13), we get

$$\beta_p^k = \frac{\sum_{ij} y_{ijk} \ (\sum_{s=1}^{p} a_s^p \ z_{ijs})}{\sum_{ij} (\sum_{s=1}^{p} a_s^p \ z_{ijs})^2}$$

$$= \frac{\sum_{ij} y_{ijk} \ [\sum_{s=1}^{p} a_s^p \ (y_{ijs} - \frac{y_{i.s}}{r} - \frac{y_{.js}}{t} + \frac{y_{..s}}{rt})]}{\sum_{ij} [\sum_{s=1}^{p} a_s^p \ (y_{ijs} - \frac{y_{i.s}}{r} - \frac{y_{.js}}{t} + \frac{y_{..s}}{rt})]^2} \quad \ldots (3.20)$$

<u>To find error sum of square</u>:

Substituting the estimators of parameters in (3.2), we get

$$R = \sum_{ijk} (y_{ijk} - \hat{\mu} - \hat{t}_i - \hat{b}_j - \hat{P}_k - \hat{t}_{ik} - \hat{b}_{jk} - \sum_{n=2}^{k} \hat{\beta}_{n-1}^{k} \ e_{ijn-1})^2$$

$$= \sum_{ijk} y_{ijk} \ (y_{ijk} - \hat{\mu} - \hat{t}_i - \hat{b}_j - \hat{P}_k - \hat{t}_{ik} - \hat{b}_{jk} - \sum_{n=2}^{k} \hat{\beta}_{n-1}^{k} e_{ijn-1})$$

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

$$\sum_{ik} \hat{t}_{ik} \ y_{i.k} - \sum_{jk} \hat{b}_{jk} \ y_{.jk} - \sum_{ijk} \sum_{n} \beta_{n-1}^{k} \ e_{ijn-1} \ y_{ijk}$$

$$= \sum_{ijk} y_{ijk}^2 - \hat{\mu} y_{...} - \sum_i (\frac{y_{i..}}{qr} - \hat{\mu}) \ y_{i..}$$

$$- \sum_j (\frac{y_{.j.}}{qt} - \hat{\mu}) \ y_{.j.} - \sum_k (\frac{y_{..k}}{rt} - \hat{\mu}) \ y_{..k} -$$

$$\sum_{ik} (\frac{y_{i.k}}{r} - \frac{y_{i..}}{qr} - \frac{y_{..k}}{tr} + \frac{y_{...}}{qrt}) \ y_{i.k}$$

$$- \sum_{jk} (\frac{y_{.jk}}{t} - \frac{y_{.j.}}{qt} - \frac{y_{..k}}{rt} + \frac{y_{...}}{qrt}) \ y_{.jk}$$

$$- \sum_{k=1}^{q} \sum_{n=2}^{k} \sum_{s=1}^{n-1} \hat{\beta}^{k}_{n-1} \; a^{n-1}_{s} \left(\sum_{ij} y_{ijk} \, y_{ijs} - \sum_{i} \frac{y_{i.s} \, y_{i.k}}{r} \right.$$

$$\left. - \sum_{j} \frac{y_{.js} \, y_{.jk}}{t} + \frac{y_{..s} \, y_{..k}}{rt} \right)$$

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

$$- \sum_{k=1}^{q} \sum_{n=2}^{k} \sum_{s=1}^{n-1} \beta^{k}_{n-1} \; a^{n-1}_{s} \left(\sum_{ij} y_{ijk} \, y_{ijs} - \sum_{i} \frac{y_{i.s} \, y_{i.k}}{r} \right.$$

$$\left. - \sum_{j} \frac{y_{.js} \, y_{.jk}}{t} + \frac{y_{..s} \, y_{..k}}{rt} \right) \quad \ldots \quad (2.21)$$

With the degrees of freedom of $q[(r-1)(t-1) - \frac{(q-1)}{2}]$

In order to derive the expressions for the sums of squares due to the effect of each factor in (3.1) a hypothesis of null differences among the different levels of each factor, which amount 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 E', contains the sum of squares due to the effect of that particular factor on which the hypothesis is made. Therefore E'-E provides the sum of squares due to that factor. In a similar manner sum of squares due to every factor in (3.1) can be derived.

To estimate treatment sum of square and testing its significance:

$H_0 : t_1 = t_2 = \ldots = t_t$

In order to test this hypothesis H_0, a model which does not have the treatment x season interaction is to be considered.

ie., we have to consider model

$$y_{ijk} = \mu + t_i + b_j + P_k + b_{jk} + \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1} + e_{ijk} \quad ..(3.22)$$

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

$$E' = \sum_{ijk} (y_{ijk} - \mu - t_i - b_j - P_k - b_{jk} - \sum_{n=2}^{k} \beta_{n-1}^{k} e_{ijn-1})^2$$

and the estimates are

$\hat{\mu} = \frac{y_{...}}{qrt}$; since it can be very well be shown that the estimation of $e_{..k}$ is zero, for every k

$$\hat{t}_i = \frac{y_{i..}}{qr} - \frac{y_{...}}{qrt} - \frac{\sum_k \sum_n \beta_{n-1}^{k} e_{i.n-1}}{qr}$$

$$\hat{b}_j = \frac{y_{.j.}}{qt} - \frac{y_{...}}{qrt} - \frac{\sum_k \sum_n \beta_{n-1}^{k} e_{.jn-1}}{qt}$$

$$\hat{P}_k = \frac{y_{..k}}{rt} - \frac{y_{...}}{qrt}$$

$$\hat{b}_{jk} = \frac{y_{.jk}}{t} - \hat{\mu} - \hat{b}_j - \hat{P}_k - \frac{\sum_n \beta_{n-1}^{k} e_{.jn-1}}{t}$$

$$= \frac{y_{.jk}}{t} - \frac{y_{.j.}}{qt} - \frac{y_{..k}}{rt} + \frac{\sum_k \sum_n \beta_{n-1}^{k} e_{.jn-1}}{qt} - \frac{\sum_{n=2}^{k} \beta_{n-1}^{k} e_{.jn-1}}{t} + \frac{y_{...}}{qrt}$$

$$\beta_p^k = \frac{1}{\sum_{ij} \hat{e}_{ijp}^2} \left[\sum_{ijk} \frac{y_{ijk}\, \hat{e}_{ijp}}{q} - \sum_i \hat{t}_i\, \hat{e}_{i.p} - \sum_j \hat{b}_j\, \hat{e}_{.jp} - \sum_{jk} \frac{\hat{b}_{jk}\, \hat{e}_{.jp}}{q} \right] \quad \ldots \quad (3.23)$$

It may be noted that the estimates of $e_{.jk}$ and $e_{.j.}$ are not zero even if we are imposing the restriction (3.4). It is also found that the estimates of t_i and b_j contains the term

$\sum_k \sum_n \beta_{n-1}^k\, e_{i.n-1}$ and $\sum_k \sum_n \beta_{n-1}^k\, e_{.jn-1}$ respectively and the estimate of β_p^k is a function of $\hat{t}_i$, $\hat{b}_j$ and $\hat{e}_{ijk}$. This makes it tedious to get an estimate of β_p^k from (3.23) independent of all other regression coefficients and e_{ijk}'s. Hence the derivation of error sum of square, E', is tedious without further assumptions or restrictions.

Under H_0, the model (3.22) reduces to

$$y_{ijk} = \mu + b_j + P_k + b_{jk} + \sum_{n=2}^{k} \bar{\beta}_{n-1}^k\, e_{ijn-1} + e_{ijk} \quad \ldots \quad (3.24)$$

The derivation of residual sum of squares from (3.24) also involves the same difficulties encountered for model (3.23).

Therefore, until the exact sums of squares attributable to the effects of factors other than the residual effect in the model (3.1) are arrived at, quantities that may approximate them can be made use of for practical purposes. The sum of squares due to systematic effects derived from the model

$$y_{ijk} = \mu + t_i + b_j + P_k + b_{jk} + t_{ik} + e_{ijk} \quad \ldots \quad (3.25)$$

which does not take the regression on residuals of yester years/seasons can very well be used for the purpose. Since the regression is made on the error terms this may not affect the result seriously. The various sum of squares from the model (3.25) along with their degrees of freedom (Das and Giri, 1986) are given below.

Treatment sum of square is given by

$$S_1^2 = \sum_i \frac{y_{i..}^2}{qr} - \left(\frac{y_{...}}{qrt}\right)^2 \quad \text{with } (t-1) \text{ degrees of freedom.}$$

The sum of squares due to seasons/years

$$S_2^2 = \sum_k \frac{y_{..k}^2}{rt} - \left(\frac{y_{...}}{qrt}\right)^2 \text{ with } (q-1) \text{ degrees of freedom.}$$

The sum of squares due to the interaction of seasons and treatments

$$S_3^2 = \sum_{ik} y_{i.k}^2 - \sum_i \frac{y_{i..}^2}{qr} - \sum_k \frac{y_{..k}^2}{rt} + \left(\frac{y_{...}}{qrt}\right)^2$$

with (q-1) (t-1) degrees of freedom. The various factors are tested against the pooled error mean square derived,using (3.21).

The F ratio for the treatment is

$$\frac{S_1^2/(t-1)}{R/q[(r-1)(t-1) - \frac{(q-1)}{2}]}$$

with degrees of freedoms of $\{(t-1), q[(r-1)(t-1)- \frac{(q-1)}{2}]\}$ and the F ratio for the treatment x season interaction is

$$\frac{S_3^2 /(q-1)(t-1)}{R/q[(r-1)\ (t-1) - \frac{(q-1)}{2}]}$$

with degrees of freedoms of $\{(q-1)(t-1), q[(r-1)\ (t-1) - \frac{(q-1)}{2}]\}$

Illustration

ILLUSTRATION

The methodology developed in chapter 3 was illustrated using data generated from 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 recorded and taken from Snedecor and Cochran (1967). They are given in Table 1.

Sum of squares due to various effects from Model (3.25) were obtained as follows:

Treatment sum of squares = 0.1781 with degrees of freedom of 2 and the treatment mean square = 0.08905

Block sum of square = 4.1499 with degrees of freedom of 20 and the block mean square = 0.207495

season x treatment interaction sum of square = 0.2105 with degrees of freedom of 6.

Season x treatment interaction mean square = 0.03508

The regression coefficients in model (3.1) were then estimated using (3.20) and are given below.

$$\hat{B} = (\hat{\beta}_i^j) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .39234 & 1 & 0 & 0 \\ .421034 & .297943 & 1 & 0 \\ .64214 & .4487 & -.01185 & 1 \end{bmatrix}$$

The error sum of squares worked out to be 1.4251 with 34 degrees of freedom. The contribution of each season to error sum of squares were also worked out and were found homogeneous when tested using Bartletts χ^2 test.

The F ratio for testing interaction mean square was 0.8369 and hence not significant. Therefore, it is customary to combine the error sum of squares and interaction sum of squares and do the testing of treatment differences against this combined mean square. Thus the new error mean square was 0.04089.

F ratio for testing treatment differences was 2.1778 which was again not significant.

The conventional analysis in the split-plot set up was carried out and is given below.

It may be noted that the treatment differences are tested against an error (a) mean square having just 10 degrees of freedom in contrast to the error mean square in the new procedure having 34 degrees of freedom. The error mean square (a) in the split-plot set up is much higher than that in the new methodology.

Analysis of variance of split-plot experiment on alfalfa

Source	df	SS	MS	F
Total	71	9.1218		
Main plots:				
Varieties	2	0.1781	0.0890	0.6534
Blocks	5	4.1499	0.8300	6.0939
Main plot error (a)	10	1.3622	0.1362	
Sub Plots:				
Date of cutting	3	1.9625	0.6542	23.3642
Date x variety	6	0.2105	0.0351	1.2535
Sub plot error	45	1.2586	0.0280	

Though it is a fact that both methods ended up with non-significant treatment differences, the new procedure is found more sensitive owing to the lower mean square error and higher degrees of freedom.

TABLE I

Date	Variety	Block 1	2	3	4	5	6	
A	Ladack	2.17	1.88	1.62	2.34	1.58	1.66	11.25
	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	
	Ladack	2.29	1.60	1.67	1.91	1.39	1.12	9.98
C	Cossack	1.86	1.70	1.81	1.54	1.67	0.88	9.46
	Ranger	1.55	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	1.33	9.66
		6.06	5.54	5.82	5.05	4.48	3.49	30.44

Discussion

DISCUSSION

Yates (1982) rightly pointed out that in split-plot experiments the subplot treatments have to be assigned randomly with in each whole plot and since the observations on the same experimental unit at different points of time have a temporal sequence, the randomisation for subplots can not be taken for granted. The random assignment of treatments is not carried out here and hence the split-plot set up is incorrect in dealing with repeated measurements.

The split-plot analysis for repeated measurements is valid when the error terms attached to the observation on the same experimental unit but taken at different time points have a constant correlation coefficient (Gill, 1988). But in experiments where the measurements have a temporal sequence, observations with a small interval of time will in general be highly correlated than those widely separated (Yates, 1982).

All these facts point to the inappropriateness of the indiscriminative use of the analysis in the split-plot set up. But in the absence of a more appropriate comprehensive analysis of data for experiments having repeated measurements, the split-plot analysis which suits to the requirements of the researcher and having comparatively less drawbacks is being widely used.

The methodology developed herein takes care of the dependence of the error terms on that of yester years/seasons in the model itself, whereas it is not contemplated in the split-plot analysis. Even from the point of sensitivity of the F test in the analysis of variance, the new methodology is superior to the split-plot analysis for testing significance of treatment differences. Since the error mean square in testing the treatment differences have more degrees of freedom in the new methodology than that in split-plot analysis, the treatment differences are estimated with more precision in the former compared to that in the latter. Error mean square for treatment difference in new methodology is likely to be much lower than that in the split-plot analysis owing to the big gap between the two error degrees of freedom as is evident from the illustration in chapter 4.

Though the exact means squares due to various factors other than error mean square could not be derived, the mean squares from model (3.24) proposed herein can be appropriately used as they are not likely to deviate far from the actuals. The deviation can only be because of the adjustments for the regression of the error terms on those in the yester years/seasons which have zero expectation.

In the illustration, the error mean square obtained using the new method is 0.04089 and have 34 degrees of freedom, whereas the error mean square (a) which is used for testing the treatment differences in split-plot set up is 0.1362 with 10 degrees of freedom. The F ratio for testing the treatment differences in the new method is 2.1738 and that in the split-plot set up is only 0.0890. Though both the procedures ended up with same conclusions the new method is superior because of the low error mean square with more degrees of freedom for error.

The difference in error mean squares and their degrees of freedom in the illustration considered herein are indicative of the sensitivity of the new methodology for comparison of treatments as compared to conventional split-plot set up.

Therefore, the new procedure developed herein may be used in preference to the split-plot analysis for analysis of data generated from repeated measurements. The proposed method is particularly very useful for experiments with perennial crops and with animals.

However the procedure can be improved up on by deriving the exact mean squares due to the systematic factors in model (3.1).

Summary

SUMMARY

A new methodology for the comprehensive analysis of data generated from experiments in which observations are repeated measurements, was developed and illustrated. The method was compared with the widely adopted split-plot analysis.

The model for the analysis of repeated measurements from the same experimental unit was introduced 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} is 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, 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 error mean square against which treatment differences are tested have more degrees of freedom in the new methodology than that in split-plot analysis.

Even from the point of sensitivity of the F test in the analysis of variance, the new methodology developed is superior to the split-plot analysis of variance for testing the significance of treatment differences.

The exact mean squares due to various factors other than error mean square could not be derived from the proposed model. Since the regression was made on the error terms of yester years/seasons and the error terms have zero expectation, the mean square for the systematic factors derived from a model without considering regression, proposed herein can be considered as a good approximation to those from the new model.

The proposed methodology was illustrated using data generated from an experiment conducted to compare three varieties of alfalfa laid out in RBD in six replications.

The error mean square for testing the treatment differences using the new method was only 0.0419147 with 34 degrees of freedom for the example considered whereas it was 0.1362 with 10 degrees of freedom in split-plot analysis. F ratio for testing treatment differences using proposed error mean square which was though not significant, was 2.1778 whereas it was only 0.6534 in split-plot analysis.

Though both the procedures ended up with same conclusion of no treatment differences, the proposed method was superior because of low error mean square with more degrees of freedom and higher F ratio.

References

REFERENCES

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

Bjornsson, H. 1987. Analysis of a series of long term grassland experiments with auto correlated error. Biometrics 34: 645-657.

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

Cole, L.W.L. and Grizzle, J.E. 1966. Application of multivariate analysis of variance to repeated measurements experiments. Biometrics. 22: 810-828.

Cullis, B.R. and McGilchrist, C.A. 1990. A model for the analysis of growth data for designed experiments. Biometrics. 46: 131-142.

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

Das,M.N. and Giri, N.C. 1986. Design and Analysis of Experiments. 2nd Edn. Wiley Estern Ltd.

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

Finney, D.J. 1956. Multivariate analysis and agricultural experiments. Biometrics. 12: 67-71.

Gill, J.L. 1986. Repeated measurements: Sensitive test for experiments with few animals. J. Anim. Sci. 63:943-954

Gill, J.L. 1988. Repeated measurements. Split-plot trend analysis versus analysis of first differences. Biometrics. 44: 289-297.

Khosla, 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.

Morrison, D.F. 1972. The analysis of a single sample of repeated measurements. Biometrics. 28: 55-71.

Patterson, 1939. Statistical Techniques in Agricultural Research. MacGraw Hill & co., New York.

Pearce, S.C. 1953. Field Experimentation with Fruit Trees and other Perennial Plants. Commonwealth Bureaux of Horticulture and Plantation Crops, Kent.

Rai, S.C. and Rao, P.P. 1980. Use of ranks in groups of experiments. J. Indian Soc. agric. Stat. 32(2): 25-33.

Rawell, J.G. and Walter, D.E. 1976. Analysing data with repeated observations on experimental unit. J. agric. Sci. , 87: 423-432.

Snedecor, G.W. and Cochran, W.G. 1967. Statistical Methods. 6th Edn., Oxford & IBH Publishing Co.

Steel, R.D.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., New York.

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

Yang, M.C.K. and Carter, R.L. 1983. One-way analysis of variance with time series data. Biometrics. 39: 747-751.

Yates, F. 1954. The analysis of experiments containing different crop rotations. Biometrics 10: 324-346.

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

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

POOLED ANALYSIS OF DEPENDENT SETS OF DATA

By

SUKUMARAN, K.

ABSTRACT OF A THESIS

Submitted in partial fulfilment of the requirement for the degree

Master of Science in Agricultural Statistics

Faculty of Agriculture
Kerala Agricultural University

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

1991

ABSTRACT

A new methodology for the analysis of data generated from experiments in which observations constitute repeated measurements from the same experimental unit at different points of time was developed. The problem of dependence of error terms in successive observations was taken care of in the model for analysis itself. The model included regression of error terms on those in the yester years/seasons. The error mean square from this model was derived using principle of least squares. The proposed method was compared with the widely adopted split-plot analysis and its superiority discussed.

The method was illustrated using data generated from an experiment conducted to compare three varieties of alfalfa laid out in RBD with six replications and observations taken in four consecutive seasons. The superiority of the new method over the split-plot analysis was evident in the example considered.