ON MISSING AND MIXED UP PLOT TECHNIQUES

By M. N. Das and P. N. Bhargava
Indian Council of Agricultural Research, New Delhi

Introduction

Experiments having missing observations are usually unavoidable. The problems due to such missing observations have been dealt with by several authors. Allen and Wishart (1930) first applied the method of fitting constants to analyse the data as non-orthogonal when there was only one missing value. This method was extended by one of the authors, Das (1954) to the case of more than one missing value in Randomised Block Design, Latin-Squares and Incomplete Block Designs. Another technique of analysis through estimation of the missing values by minimising the error sum of squares obtained by substituting an unknowns for the missing observations was introduced by Yates (1933). The third method consists in introducing pseudo-variates and applying the technique of covariance by inserting zero for the missing observations and was suggested by Bartlett (1937).

The second method due to Yates is widely used and the others are usually considered lengthy. Das (1954) showed that for more than one missing plot in Randomised Block and Latin-Square Designs the method of fitting constants can in many cases be profitably applied.

The problem of mixed up observation was first encountered by Bose and Mahalanobis (1938) and they solved the problem by modifying the method of Yates. Later Nair (1940) suggested a solution to the problem of mixed up plots by applying the technique of covariance introduced by Bartlett.

In the present paper an attempt has been made to examine the interrelationship between the last two techniques with an examination of the rationale behind the covariance technique.

For mixed up observations an alternative definition of concomitant variates has been given and a solution for more than one group of mixed up observations has also been obtained.

Rationale behind the Technique

With one missing observation substituted by x_1, the position of the affected plot remaining unknown, the data which are now of two types, *viz.*, (i) those which are the observations of a random variate y_i and (ii) an unknown x_1 not belonging to the population of the observation, can be represented for the i-th plot by the variate $Z_i = y_i + x_1 \cdot x_{1i}$, if y's are the available observations and x_{1i} a pseudo-variate such that it takes the value zero in all unaffected plots and in the case of the affected plot it takes the value unity and y_i, zero.

What is actually wanted is the analysis of the variate Z_i. It is evident from the relation, $Z_i = y_i + x_1 \cdot x_{1i}$ that the results of analysis of variance of Z_i can be obtained from those of y_i adjusted for x_{1i}. The relation shows that x_1 is the regression coefficient of y_i on x_{1i} with the sign changed. In case of two missing plots with x_1 and x_2 as their substitutes, the data for the i-th plot can be represented by $Z_i = y_i + x_1 . x_{1i} + x_2 \cdot x_{2i}$; if x_{1i} and x_{2i} are defined as before and y_i takes the value zero in each of the affected plots. This way of representation can be extended for any number of affected plots and it remains true whatever the design. The covariance technique can thus be applied for any design having any number of missing observations and bears an exact correspondence with the method of substitution by Yates.

In the case of mixed up observations affecting only two plots, say, M, it is the total of the observations from these plots, the data can be completed as in the case of missing plots, by substituting $(M/2) + x_1$ and $(M/2) - x_1$ in the two affected plots where x is an unknown which with $M/2$ gives the estimated yield of one of the plots. The data, thus completed, can for the i-th plot be represented by $Z_i = y_i + x_1 . x_{1i}$ if x_{1i} takes the value zero in all unaffected plots and in the affected plots it takes the value 1 in one plot and -1 in the other, while y_i takes the value $M/2$ in each of these plots. Hence, x_1 is the regression coefficient of y_i on x_{1i} with its sign changed. If there be K plots mixed up to give the total M, the data for the i-th plots can be represented similarly by $Z_i = y_i + x_1 \cdot x_{1i} + x_2 \cdot x_{2i} + \ldots + x_{K-1} \cdot x_{(K-1)i}$ if each of the variates x_{ji} ($j = 1, 2, \ldots K - 1$) take values zero for all unaffected plots and in one of the affected plots each of the variates x_{ji} takes the value -1 and in the remaining $(K-1)$ plots, the $(K-1)$ variates take the value as in the following scheme, the value of y_i being taken as M/K in each of the affected plots.

The scheme of pseudo-variate values suggested by Nair (1940) for the same purpose is different and somewhat complicated. The

Scheme showing the Values for the Pseudovariate and Observed Variate for the affected plots in the case of mixed up plots

Plot number \ Pseudo variate	1	2	3	. . .	$k-1$	Value of the observed variate y for the affected plots
1 . .	-1	-1	-1	. . .	-1	M/k
2 . .	1	0	0	. . .	0	M/k
3 . .	0	1	0	. . .	0	M/k
. . .						.
. . .						.
. . .						.
. . .						.
. . .						.
k . .	0	0	0	. . .	1	M/k

scheme given here has come out as just an extension of the scheme necessary with $(K-1)$ missing observations and actually corresponds to the following substitution of unknowns in the affected plots.

1st plot $\frac{M}{K} - x_1 - x_2 - x_3 \ldots - x_{K-1}$.

2nd plot $\frac{M}{K} + x_1$

3rd plot $\frac{M}{K} + x_2$

. . . .

. . . .

. . . .

Kth plot $\frac{M}{K} + x_{K-1}$

So that by adding all the substitute for y the affected plots we get M. Thus, it is the nature of the unknown substitute which determines the pseduo-variates. If again there be more than one mixed up total say two, *viz.*, M_1 and M_2 such that M_1 is based on K_1 observations and M_2 on K_2, then the data for the i-th plot can be represented by:

$$Z_i = y_i + x_1 \cdot x_{1i} + \ldots + x_{K_1-1} \cdot x_{(K_1-1)i}$$
$$+ x_1' \cdot x'_{1i} + \ldots + x'_{K_2-1} x' \cdot_{(K_2-1)i}.$$

If y_i takes the values in the plots as M_1/K_1 for each of the plots giving M_1 and M_2/K_2 for each plot giving M_2 and x_{ji} and x_{ji}' take the value zero in all unaffected plots and in the affected plots they take the values independently as described in the case of single total based on K mixed up observations. The variates x_{ji} corresponding to the first set of affected plots will take the value zero in all plots not in the first set and *vice versa*.

In this way the technique can be extended to cover any number of mixed up totals.

ADVANTAGES OF COVARIANCE TECHNIQUE

With the increase in the number of missing values in complex designs like asymmetrical confounded factorial designs, the procedure suggested by Yates becomes very laborious, if the iterative method of solution of normal equations be adopted, as it involves repeated calculation of several estimates from very complex expressions. In covariance technique this can be avoided by forming the equations once for all and solving them by any of the existing methods. Of course, in the case of Yates procedure, also the normal equations can be actually written down by collecting terms containing the first two powers of the missing values substituted in the error s.s., but it appears that the formation of normal equations from Yates approach is somewhat indirect, and not so suitable for mechanisation and less easily understandable than when they are formed as multiple regression equations.

Though the estimates of missing values are available from both the methods, Yates technique cannot easily supply some of the results like S.E. of treatment differences and also the adjusted treatment S.S.

SUMMARY

An investigation of the interrelationship among the different methods of analysis of experiments having missing or mixed up plots, has been made. The rationale behind the technique of co-variance analysis of experiments with missing observations has been discussed. The technique of covariance analysis has been extended in the case of mixed up observations to cover more than one group of mixed up totals. A more convenient set of pseudo-variates than what has been suggested by Nair for solving problems of mixed up total, has been obtained.

ACKNOWLEDGEMENT

We are grateful to Dr. V. G. Panse for the opportunity he made available to us as also for his encouragement which contributed a good deal for the success of the paper.

REFERENCES

1. Allan, F. E. and Wishart — *J. Agric. Sci.*, 1930, **20**, 399–406.
2. Bartlett, M. S. .. *J. Roy. Stat. Soc. Suppl.*, 1937, **4**, 137–83.
3. Das, M. N. .. *J. Ind. Soc. Agric. Stat.*, 1953, **5**, 161–78.
4. ———— .. *Ibid.* 1956, **8**, 76–83.
5. Nair, K. R. .. *Sankhya*, 1940, **4**, 581.
6. Yates, F. .. *Emp. J. Exp. Agric.*, 1933, **1**, 129–42.