ANALYSIS OF COVARIANCE IN INCOMPLETE BLOCK DESIGNS WITH OR WITHOUT MISSING PLOTS

By M. N. Das

Indian Council of Agricultural Research, New Delhi

Introduction

Though Cornish (1940) discussed the method of analysis of covariance in quasifactorial designs, it appears that for the general incomplete block designs, no simplified method of analysis of covariance is available in literature. A general method of analysis of covariance of non-orthogonal data was given by the author (Das, 1953) but this method is not based on, nor does it utilise the results of analysis of variance of each of the variates, as is done in the case of orthogonal data. As such it is not suitable for the analysis of data from the various incomplete block designs involving two-way classification, where the results of analysis of variance are readily available, though its utility is great for the analysis of general non-orthogonal data where the results of analysis of variance is not readily available. An attempt has thus been made in this paper to give a simplified method based on the results obtained from the analysis of variance of each of the variates, suitable for the different incomplete block designs both with or without missing plots including randomised block designs with missing plots.

Method

Taking the model $y_{ij} = \mu + t_i' + b_j + ax_{ij} + \epsilon_{ij}$, where μ is a constant, t_i', the effect of the ith treatment, b_j, that of the jth block, a, some constant depending on the linear relation of y with x and ϵ_{ij}, a random variable with zero mean and a constant variance, σ^2, it has been shown by the author (Das, 1953) that the best estimates of t_i''s and 'a' after eliminating the effects of the blocks, are obtainable from

$$Q_i = (r_i - \lambda_{ii})\, t_i' - \sum_{m \neq i} \lambda_{im} t_m' + aQ_{i(x)} \tag{1}$$

and

$$a = d - \frac{\Sigma\, t_i' Q_{i(x)}}{\alpha}, \tag{2}$$

where λ_{im}'s are constants depending on the parameters of the design, $d = \beta/\alpha$, β being the 'within block sum of products' and α, the 'within block S.S. for x', Q_i is the total for the y variate of the ith treatment less the sum of averages of those blocks in which the ith treatment occurs, $Q_{i(x)}$ denotes similar total for the x-variate and r_i is the replication of the ith treatment.

From (1)

$$(r_i - \lambda_{ii})\, t_i' - \sum_{m \neq i} \lambda_{im} t_m' = P_i \tag{3}$$

where

$$P_i = Q_i - aQ_{i(x)}.$$

The solution of these equations can be obtained by replacing Q_i's in the solution of t_i, *i.e.*, for the y-variate alone, by the P_i's.

Thus if the solution for t_i taking the y-variate only be

$$t_i = \Sigma\, C_{ij} Q_j,$$

then the solution of (3) is given by

$$t_i' = \Sigma C_{ij} P_j = t_i - at_{i(x)}$$

where t_i and $t_{i(x)}$ are respectively the estimates of the ith treatment from the y- and x-variates. But unless an estimate of 'a' is available t_i' cannot be obtained.

From (2) we get after substituting for the value of t_i'

$$a = d - \frac{\Sigma\, Q_{i(x)}\, \{t_i - at_{i(x)}\}}{\alpha}$$

whence

$$a = \frac{\beta - \Sigma\, t_i Q_{i(x)}}{\alpha - \Sigma\, t_{i(x)} Q_{i(x)}}.$$

'a' is thus estimated by the regression coëfficient obtained from the error line in the analysis of variance-covariance table of the design concerned. Evidently the sum of products in this line is obtained from $\{\beta - \Sigma t_i Q_{i(x)}\}$ which is also equal to $\{\beta - \Sigma t_{i(x)} Q_i\}$.

The error sum of squares adjusted for the x-variate is obtained from

$$\Sigma\, (y_{ij} - \mu - t_i' - b_j - ax_{ij})^2,$$

Eliminating b_j's with the help of the normal equations this becomes

$\Sigma\, y_{ij}^2 - \Sigma\, t_i' Q_i - \Sigma\, \beta_j^2/k_j - a\beta$, where k_j is the size of the jth block.

Substituting for t_i' it can be shown easily that

$$\text{Adjusted Error S.S.} = \text{Error S.S. for } y - \frac{\{\beta - \Sigma\, t_i Q_{i(x)}\}^2}{\{a - \Sigma\, t_{i(x)} Q_{i(x)}\}}$$

Thus the adjusted error sum of squares can be obtained from the error line of the analysis of variance-covariance table of any design exactly in the same manner as in orthogonal designs.

The adjusted treatment S.S. can now be obtained from:

$$\text{Adjusted Treatment S.S.} = \text{Within block S.S. for } y - d\beta$$
$$- \text{Adjusted error S.S.}$$

As (within block S.S. for $y - d\beta$) is the adjusted within block S.S., and the total of sum of squares or products due to treatments and error is equal to within block sum of squares or products, it is clear from above that the adjusted treatment sum of squares also can be obtained in the same manner as in orthogonal designs. The sum of products in the treatment line is evidently given by $\Sigma\, t_i Q_{i(x)}$ or $\Sigma\, t_{i(x)} Q_i$.

The variance of the difference between any two treatments $(t'_i - t_m')$ say can be obtained, as usual, by collecting the coefficients of Q_i and Q_m in $(t_i' - t_m')$ and subtracting them. Thus taking V to stand for the variance of $(t_i - t_m)$, we get

$$\text{Var}\,(t_i' - t_m') = V + \frac{\{t_{i(x)} - t_{m(x)}\}^2}{a - \Sigma\, Q_{i(x)} t_{i(x)}}\,\sigma^2$$

where σ^2 (which is also contained in V) is the adjusted error mean square.

Though these results have been obtained without reference to any particular design, it is evident that the results of analysis of covariance of any design with two-way classification can be obtained from them once the solutions of the normal equations are available for each of the variates.

In cases of missing plots the method of analysis is valid when both x and y values from one or more plots are missing. Also in such cases, this method can be applied only if the solutions of t_i are available. Expressions giving such solution have been given by the author (Das, 1954) for randomised block designs with balanced incompleteness. In cases of missing plots in partially balanced and other incomplete block designs expressions for t_i's have been given by the author (Das, 1955) when there are one or two missing plots as also when a block is partly missing.

Though the method has been obtained for designs with two-way classification, results for Youden Square when there is no missing value can easily be obtained as the rows are orthogonol to blocks and treatments. The different results can be obtained by replacing β by β-Row.S.P., α by (α-Row.S.S.) etc., for x and within block S.S. for y by within block S.S. for y-Row sum of squares for y.

AN EXAMPLE

The method has been illustrated by analysing a partially balanced incomplete block design with two associate classes having two missing values.

The y-variate values are the same as utilised by the author (Das, 1955) and were obtained in the following manner:

Eight numbers, *viz.*, 4, 5, 7, 9, 10, 12, 13 and 15 were taken to represent the effects of eight treatments V_1, V_2, V_3, V_4, $\cdots$, V_8 in order. Each of the numbers was repeated 5 times to give in all 40 numbers. These 40 numbers were then made into 8 groups of 5 each so as to give the contents of the 8 blocks of the partially balanced incomplete block design with $b = v = 8$, $r = k = 5$, $\lambda_1 = 4$, $\lambda_2 = 2$, $n_1 = 3$, $n_2 = 4$ and $p'_{11} = 2$. Next another set of eight numbers, *viz.*, 5, 2, 3, 8, 7, 1, 5 and 4 were taken to represent the block effects and the first of these numbers was added to each of the five numbers in the first group forming the first block; the second number was similarly added to each of the numbers of the second block and so on for other blocks so that the 40 totals thus obtained represented the sum total of treatment and block effects. Afterwards 40 numbers lying between 3 and -4 were chosen at random and added one to each of the 40 totals previously obtained. The numbers thus obtained formed the data and are shown below together with the values of the x-variate which have been obtained from the above simply by subtracting from each of the numbers, its treatment effect and then adding one to each of the 40 remainders to avoid negative sign.

The treatment number has been given in bracket beside the data.

BLOCKS

(1)	x	*4*	(1)	6	(2)	9	(3)	5	(4)	5	(5)
	y	*7* m		10		15		13		14	
(2)	x	4	(1)	4	(2)	2	(3)	1	(4)	*8*	(6)
	y	7		8		8		9		*14* n	
(3)	x	7	(1)	1	(2)	6	(3)	4	(4)	5	(7)
	y	10		5		12		12		17	

BLOCKS

(4)	x	8	(1)	5	(2)	11	(3)	12	(4)	8	(8)
	y	11		9		17		20		22	
(5)	x	6	(1)	4	(5)	9	(6)	8	(7)	7	(8)
	y	9		13		20		20		21	
(6)	x	5	(2)	0	(5)	3	(6)	3	(7)	3	(8)
	y	9		9		14		15		17	
(7)	x	8	(3)	3	(5)	7	(6)	6	(7)	7	(8)
	y	14		12		18		18		21	
(8)	x	4	(4)	5	(5)	5	(6)	5	(7)	7	(8)
	y	12		14		16		17		21	

The first plot in the first block and the last plot in the second block which are marked by m and n respectively, have been taken to be missing. The expected values in these plots are $m_x = 6$, $n_x = 3$, $m_y = 9$ and $n_y = 14$.

For the analysis it is first necessary to get estimates of t_i and $t_{i(x)}$. A brief account of the method of estimating t_i or $t_{i(x)}$ obtained by the author (Das) is given below.

If we define a variate m_i (which may be called treatment adjustment for the missing plot of m) such that it takes the value $1 - 1/k$ for the treatment in the plot of m, $-1/k$ for each of the other treatments in the block of m and zero for all other treatments and a similar variate, n_i for the missing plot of n, then taking m and n as the substitutes of the missing values to be estimated by minimising the error variance, it has been shown that

$$t_i = t_i'' + mM_i + nN_i$$

where

$$t_i'' = c_1 Q_i' + c_2 \Sigma Q'_{i1}$$

i.e., the solution of t_i from the normal equations formed by taking zero for the missing values which again implies that Q_i' is the adjusted total for the ith treatment obtained by taking zero for the missing values; $\Sigma Q'_{i1}$ is the summation over those Q_i''s which are the first associates of the ith treatment;

$$M_i = c_1 m_i + c_2 \Sigma m_{i1}$$

and

$$N_i = c_1 n_i + c_2 \Sigma n_{i1},$$

Σm_{i1} and Σn_{i1} having similar meaning as $\Sigma Q_{i1}'$ and the estimates of m and n can be obtained from

$$m = \frac{\left(\frac{k-1}{k} - \Sigma n_i N_i\right)\left(\frac{B_m}{k} + \Sigma M_i Q_i'\right) + \left(\frac{\delta}{k} + \Sigma M_i n_i\right)\left(\frac{B_n}{k} + \Sigma N_i Q_i'\right)}{\left(\frac{k-1}{k} - \Sigma m_i M_i\right)\left(\frac{k-1}{k} - \Sigma n_i N_i\right) - \left(\frac{\delta}{k} + \Sigma M_i n_i\right)^2}$$

and

$$n = \frac{\left(\frac{k-1}{k} - \Sigma m_i M_i\right)\left(\frac{B_n}{k} + \Sigma N_i Q_i'\right) + \left(\frac{\delta}{k} + \Sigma M_i n_i\right)\left(\frac{B_m}{k} + \Sigma M_i Q_i'\right)}{\left(\frac{k-1}{k} - \Sigma m_i M_i\right)\left(\frac{k-1}{k} - \Sigma n_i N_i\right) - \left(\frac{\delta}{k} + \Sigma M_i n_i\right)^2}$$

where B_m is the total of the block with the plot of m taking the missing values as zero, B_n is the similar total of the block containing n and $\delta = 1$ or 0 according as both the missing plots are in the same block or not.

As c_1 and c_2 are the coefficients of Q_i and ΣQ_{i1}, respectively in the solution of t_i in P.B.I.B. design, their values have been obtained first from the parameters of the design following Bose and Nair (1939). These have been obtained for the design as $c_1 = 45/192$ and $c_2 = 5/192$.

The other expressions required for the analysis can be obtained from the following table:—

Treatment Nos.	Q_i'	$Q_{i(x)}'$	m_i	n_i	Σm_{i1}	Σn_{i1}	$\Sigma Q_{i1}'$	$\Sigma Q_{i(x)1}'$	$t_i'' \times 192$	$t_{i(x)}'' \times 192$	$M_i \times 192$	$N_i \times 192$
1	−23·4	−2·4	·8	−·2	−·6	−·6	−3·8	+7·0	−1072	−73	33	−12
2	−15·6	−2·4	−·2	−·2	·4	−·6	−11·6	7·0	−760	−73	−7	−12
3	5·6	9·2	−·2	−·2	·4	−·6	−32·8	−4·6	88	391	−7	−12
4	6·2	0·2	−·2	−·2	·4	−·6	−33·4	4·4	112	31	−7	−12
5	−10·4	−9·0	−·2	0	0	·8	37·6	4·4	−280	−283	−9	4
6	−0·4	0·8	0	·8	−·2	0	27·6	−5·4	120	9	−1	36
7	13·8	1·4	0	0	−·2	·8	13·4	−6·0	688	33	−1	4
8	24·2	2·2	0	0	−·2	·8	3·0	−6·8	1104	65	−1	4

As checks the total of each column should be zero. From this table

$$\Sigma M_i Q_i' = \Sigma m_i t_i'' = -689 \cdot 6/192$$
$$\Sigma N_i Q_i' = \Sigma n_i t_i'' = 422 \cdot 4/192$$
$$\Sigma M_i Q_{i(x)}' = \Sigma m_i t_{i(x)}'' = -51 \cdot 6/192$$
$$\Sigma N_i Q_{i(x)}' = \Sigma n_i t_{i(x)}'' = -48 \cdot 0/192$$
$$\Sigma n_i M_i = \Sigma m_i N_i = -3 \cdot 2/192$$
$$\Sigma m_i M_i = 32 \cdot 4/192 \text{ and } \Sigma n_i N_i = 38 \cdot 4/192.$$

Substituting these values in the expressions for m and n,

$$m_x = 7 \cdot 41, \quad m_y = 10 \cdot 41$$
$$n_x = 3 \cdot 04, \quad n_y = 14 \cdot 04$$

The values of t_i and $t_{i(x)}$ have now been obtained as shown below:

Treatments	$t_i \times 192$	$t_{i(x)} \times 192$	Q_i''	$Q_{i(x)}''$
1	− 896·95	135·16	−14·6	2·05
2	−1001·35	−161·57	−19·8	− 4·20
3	− 153·35	302·57	1·4	7·40
4	− 129·35	− 57·43	2·0	− 1·60
5	− 317·53	−437·56	−13·0	−10·25
6	615·03	111·17	6·0	3·00
7	733·75	37·76	13·8	1·40
8	1149·75	69·76	24·2	2·20

The last two columns give the adjusted treatment totals for each of the two variates obtained from the incomplete data taking the missing values as missing and not zero as in Q_i' or $Q_{i(x)}'$.

The adjusted treatment S.S. and S.P. have then been obtained as

Adjusted treatment S.S. for $y = \Sigma t_i Q_i'' = 407 \cdot 38$

Adjusted treatment S.S. for $x = \Sigma t_{i(x)} Q_{i(x)}'' = 43 \cdot 29$

Adjusted treatment S.P. $= \Sigma t_i Q_{i(x)}'' = \Sigma t_{i(x)} Q_i'' = 52 \cdot 59.$

The 'within block S.S. and S.P.' obtained from the incomplete data together with the adjusted treatment and error S.S. and S.P. are shown below:—

	Σy^2	Σxy	Σx^2	Adj. S.S. (y)
Within block (From incomplete data)	480·80	126·00	116·70	..
Treatment (Adjusted)	407·38	52·59	43·29	344·75
Error (By subtraction)	73·42	73·41	73·41	·01

As it should be, the adjusted error S.S. has come out to be zero and the value of 'a' unity.

The variance of the difference between any two treatments can now be obtained easily with the help of $t_{i(x)}$'s and V, the variance of the difference in analysis of variance. Expressions for V have been given by the author (Das,[4]) for the various cases.

SUMMARY

A general method of intra-block analysis of covariance based on the results of analysis of variance and suitable for all incomplete block designs with two-way classification, has been given. The results cover the cases of incompleteness due to missing values in any design including randomised block. An expression for finding the standard error of the difference between any two adjusted treatment effects has been deduced for the general case. The method of analysis has been illustrated by analysing a partially balanced incomplete block design with two missing values.

REFERENCES

1. Bose, R. C. and Nair, K. R. .. *Sankhya*, 1939, **4**, 337.
2. Cornish, E. A. .. *Ann. Eugen. Lond.*, 1940, **10**, 269.
3. Das, M. N. .. *J. Ind. Soc. Agric. Stat.*, 1953, **5**, 161.
4. ——— .. *Ibid.*, 1954, **6**, 58.
5. ——— .. *Ibid.*, 1955, **7**, 111.
6. Kempthrone, O. .. *The Design and Analysis of Experiments*, John Wiley and Sons, London, 1952, p. 558.