

# A GENERALISED BALANCED DESIGN

BY M. N. DAS

*Indian Council of Agricultural Research, New Delhi*

## INTRODUCTION

IN field plot experiments involving a large number of treatments, it becomes necessary to take recourse to incomplete block designs in order to reduce the heterogeneity within blocks. On the other hand, in experiments involving single plant progenies with variable amounts of seed (Hutchinson and Panse, 1937) or a small number of treatments with animals as experimental units, it becomes often desirable to adopt super-complete block designs in order to avoid wastage of seed or animals. Thus, in an experiment involving four treatments, if litters of six animals each are available, the usual procedure is to take only four animals from each litter and reject the remaining two. If, instead, the experiment is suitably designed so as to utilise all the six animals, it not only saves wastage but also increases the precision. The main objection to such a procedure in the case of agricultural experiments is due to the increase in block size and consequent loss of precision which is inversely related to block size. In experiments with animals as experimental units on the other hand, the heterogeneity within litters does not most likely depend on the size of the litter. Such consideration also leads to the necessity of having designs by means of which litters of different sizes can be utilised in the same experiment. In field plot experiments, this actually corresponds to designs with unequal blocks.

Keeping all such considerations in view, a generalised balanced design has been defined in this paper with provision of replications within 'block', if necessary. The usual complete block designs, the balanced incomplete block designs with or without provision of some extra treatments which are present in every block and various types of super-complete block designs come out as particular cases of this design.

Pearce (1953) reported an interesting experiment, which was adopted in East Malling Fruit Research Station to meet an *ad hoc* situation. It was first planned as a randomised block design with four blocks of seven plots each to accommodate five weed-killers and two control treatments. Through some unforeseen circumstances, one of the weed-killers could not be obtained though plots were kept

ready for it. What was done on the spur of the moment was to replace the absent treatment by each of the available ones in turn in the different blocks. It became thus a super-complete non-orthogonal design. Pearce stated that he was not aware if the method of analysis of such a design is available but guessed that perhaps the method of analysis of the general incomplete block designs given by Rao (1947) might cover such designs. Rao's method, however, does not cover designs with replications within blocks. Actually, the above 'Nonce design', as it has been named by Pearce, is a particular case of the general design defined here.

The method of intra-block analysis of the general design together with expressions for the variances of treatment contrasts has also been presented in the paper.

## 2. DEFINITION OF THE DESIGN

Taking a block to mean a group of experimental units like plots, animals, etc., having some common features, the design has been defined as below:—

If in the  $j$ th of  $b$  blocks, the  $i$ th treatments occurs  $n_{ij}$  times, such that for  $q$  of the  $t$  treatments,  $n_{ij}$  is constant in all the blocks, being equal to  $n \geq 0$ , while for the other  $(t - q)$  treatments it takes two types of values in the different blocks, viz.,  $s$  and  $s + p$  ( $s \geq 0, s + p \geq 0$ ) such that the cells in the treatment  $(t - q) \times$  blocks ( $b$ ) table taking the frequencies  $(s + p)$  give rise to a balanced incomplete block design with  $(t - q)$  treatments,  $b$  blocks and some  $\lambda \geq 0$ . To these  $b$  blocks are to be added another  $b'$  blocks such that the frequencies  $n_{ij} \geq 0$  are the same within the same block though they may differ from block to block.

Such a design involving  $t$  treatments and  $b + b'$  blocks has been called a generalised balanced design. Taking the parameters of the balanced incomplete block design as  $b = b, v = t - q, r, \kappa, \lambda$ , the cell frequencies and the marginal total frequencies, the latter giving replications and block sizes of the design, have been shown in tabular form on the next page.

When  $n_i = p = 0$  and  $n = s = 1$ , it becomes the ordinary Randomised block design. Again, when  $n_i = n = s = 0$  and  $p = 1$ , it becomes the balanced incomplete block design. A balanced incomplete block design with some extra treatments present in every block is obtainable by putting  $n = 1 = p$  and  $s = 0$ .

Blocks	Treatments						Total			
	1	2	..	q,	q+1,	q+2		.....	t	
1	n	n	..	n	s	s+p	....	s	$K=nq+s(t-q)+kp$	
2	n	n	..	n	s+p	s	....	s+p	K	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
b	n	n	..	n	s	s+p	....	s+p	K	
b+1	n <sub>1</sub>	n <sub>1</sub>	..	n <sub>1</sub>	n <sub>1</sub>	n <sub>1</sub>	....	n <sub>1</sub>	tn <sub>1</sub>	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
b+b'	n <sub>b'</sub>	n <sub>b'</sub>	..	n <sub>b'</sub>	n <sub>b'</sub>	n <sub>b'</sub>	....	n <sub>b'</sub>	tn <sub>b'</sub>	
Total	$nb+\sum n_i$		R <sub>1</sub>	..	R <sub>1</sub>	$sb+rp$	R <sub>2</sub>	....	R <sub>2</sub>	$+\sum n_i \equiv R_2$

The design reported by Pearce (1953) has the cell frequencies as shown below:—

Blocks	Treatments				
	0	A	B	C	D
1	2	2	1	1	1
2	2	1	2	1	1
3	2	1	1	2	1
4	2	1	1	1	2

This evidently is a particular case of the general design. We have here  $q = 1, t = 5, n = 2, s = 1$  and  $p = 1$  and  $n_i = 0$ , such that the cells with frequency 2 under the treatments A, B, C and D form the balanced incomplete block design with parameters  $v = b = 4, r = \kappa = 1, \lambda = 0$ . As the design provides for replications within blocks, an estimate of

the error within cells can be obtained and hence the interaction between blocks and treatments can be tested.

### 3. METHOD OF ANALYSIS

If  $y_{ijk}$  denotes the  $k$ th observation from the  $i$ th treatment in the  $j$ th block, then taking the usual model,  $y_{ijk} = \mu + t_i + b_j + \epsilon_{ijk}$  where  $\mu$ ,  $t_i$  and  $b_j$  are constants denoting the general mean, the effect of the  $i$ th treatment and that of the  $j$ th block respectively and  $\epsilon_{ijk}$ , a random variable with zero mean and a constant variance,  $\sigma^2$ , it has been shown by the author (Das, 1953) that the normal equations involving the treatment effects alone after eliminating  $b_j$ 's and one of the treatments say, the  $p$ th, with the help of the restriction  $\sum t = 0$ , come out to be

$$N_i.t_i - \sum_k t_k \sum_j \frac{n_{ij} (n_{ki} - n_{pj})}{n_j} = Q_i \begin{pmatrix} i = 1, 2, \dots, (t-1) \\ k = 1, 2, \dots, (t-1) \\ j = 1, 2, \dots, b \end{pmatrix}$$

where  $N_i$  is the replication of the  $i$ th treatment,  $n_j$ , the size of the  $j$ th block and  $Q_i$  is the adjusted total of the  $i$ th treatment and is given by  $Q_i = T_i - \sum n_{ij} B_j/n_j$ ,  $T_i$  being the total of the  $i$ th treatment and  $B_j$  that of the  $j$ th block.

In the present design there are two types of replications of the treatments, viz.,  $R_1 = (bn + n_1 + n_2 + \dots + n_r)$  or  $R_1 = nb + \sum n_i$  for the first  $q$  treatments, to be called hereafter the first set of treatments, and  $R_2 = sb + rp + \sum n_i$ , i.e.,  $R_2 = R + \sum n_i$  where  $R = sb + rp$  for the other  $(t - q)$  treatments involved in the B.I.B. design, to be called hereafter the second set of treatments. The block size of each of the first  $b$  blocks forming the B.I.B. design is  $nq + s(t - q) + \kappa p = K$ , say, the size of the other blocks being different in each case as shown in the margin of the table. Taking the  $p$ th treatment, which has been eliminated, to be one in the first set of  $q$  treatments so that  $n_{pi} = n$  or  $nl$  for all  $j$ , the expression

$$\sum \frac{n_{ij} (n_{ki} - n_{pj})}{n_j} \text{ reduces to}$$

$$\begin{aligned} & \frac{1}{K} \{ \lambda (s + p) (s + p - n) + (b - 2r + \lambda) s (s - n) \\ & \quad + (r - \lambda) (s - n) (s + p) + (r - \lambda) s (s + p - n) \} \\ & = \frac{R (s - n) + p (\lambda p + sr)}{K} \end{aligned}$$

for all  $\kappa \neq i$ , both  $i$  and  $\kappa$  standing for treatments in the second set.

The above expression can be obtained easily, from a consideration of the fact that for the two treatments  $i$  and  $k$  both  $n_{ij}$  and  $n_{ki}$  can take values (i)  $(s + p)$  just in  $\lambda$  blocks and (ii)  $s$  in  $b - 2r + \lambda$  blocks; in  $(r - \lambda)$  blocks one of them takes the value  $s + p$  and the other  $s$  and *vice versa* in the remaining  $(r - \lambda)$  blocks.

When  $i = \kappa$ , the expression takes the value

$$\begin{aligned} & \frac{1}{K} \{r(s + p)(s + p - n) + (b - r)s(s - n)\} \\ &= \frac{R(s - n) + rp(s + p)}{K} \end{aligned}$$

As  $n_{kj} = n_{pj} = n$  for treatments in the first set, the normal equations corresponding to the treatments in the second set, will not contain any treatment in the first set, as each of them enters these equations with coefficients involving  $(n_{kj} - n_{pj})$  which is zero.

Thus, the normal equations corresponding to the treatments in the second set can be written as

$$\begin{aligned} & \left\{ R_2 - \frac{(s - n)R + rp(s + p)}{K} \right\} t_m \\ & - \frac{R(s - n) + p(\lambda p + sr)}{K} \sum_{m' \neq m} t_{m'} = Q_m \end{aligned} \quad (1)$$

where  $t_m$  and  $t_{m'}$  stand for treatments in the second set.

The normal equations for the treatments in the first set come out to be

$$R_1 t_i - \frac{nR - bn^2}{K} \sum t_m = Q_i \quad (2)$$

where  $t_i$  stands for treatments in the first set.

Thus, once the solution of the equations corresponding to the treatments in the second set is available, those for the treatments in the first set can be easily obtained.

The equations in (1) can also be written as

$$\begin{aligned} & \left\{ R_2 - \frac{p^2(r - \lambda)}{K} \right\} t_m \\ & - \frac{R(s - n) + p(\lambda p + sr)}{K} \sum t_m = Q_m \end{aligned}$$

Summing it over  $m$  we get

$$\sum t_m \left[ \left\{ R_2 - \frac{p^2 (r - \lambda)}{K} \right\} - (t - q) \left\{ \frac{R(s - n) + p(\lambda p + rs)}{K} \right\} \right] = \sum Q_m.$$

Putting

$$\left\{ R_2 - \frac{p^2 (r - \lambda)}{K} \right\} \equiv A$$

and

$$\frac{R(s - n) + p(\lambda p + rs)}{K} \equiv B,$$

we get

$$\sum t_m = \frac{\sum Q_m}{A - (t - q) B}.$$

and

$$t_m = \frac{1}{A} \left\{ Q_m + \frac{B \sum Q_m}{A - (t - q) B} \right\}.$$

Hence

$$t_i = \frac{1}{R_1} \left[ Q_i + \frac{n(R - bn) \sum Q_m}{K \{ A - (t - q) B \}} \right].$$

The treatment  $t_p$  which was eliminated can also be obtained from the above expression simply by replacing  $Q_i$  by  $Q_p$ .

The solution for particular designs can be obtained by first finding the parameters of the design and then substituting them in the above expressions. Thus, for Pearce's 'Nonce design', we have  $t = 5$ ,  $q = 1$ ,  $n = 2$ ,  $s = 1$ ,  $p = 1$ ,  $b = v = 4$ ,  $r = \kappa = 1$ ,  $\lambda = 0$ ,  $K = 7$ ,  $R_1 = 8$ ,  $R_2 = R = 5$ , and  $n_i = 0$ .

Hence

$$A = 5 - \frac{1}{7} = \frac{34}{7}.$$

$$B = \frac{-5 + 1}{7} = \frac{-4}{7}.$$

whence

$$t_m = \frac{7}{34} \left( Q_m + \frac{2}{25} Q_0 \right)$$

and

$$t_0 = \frac{7}{50} Q_0.$$

where  $t_0$  stands for the control and  $t_m$  for the other treatments.

Once the solution of the treatment effects has been obtained, the treatment sum of squares can be obtained as usual from  $\Sigma tQ$  and the S.S. due to all the fitted constants, from  $\Sigma B_j^2/n_j + \Sigma tQ$ . By subtracting the sum of squares due to the fitted constants from  $\Sigma T_{ij}^2/n_{ij}$  where  $T_{ij}$  is the total of the  $i$ th treatment as obtained from the  $j$ th block, the interaction sum of squares can be obtained. The within block pure error sum of squares can be obtained, as usual, from the within cell sum of squares.

By expressing any treatment contrast as a linear function of their adjusted totals, the variance of the difference between any two treatments can be obtained easily.

Thus

$$\text{Variance } (t_m - t_{m'}) = \frac{2\sigma^2}{A}$$

$$\text{Variance } (t_i - t_{i'}) = \frac{2\sigma^2}{R_1}$$

$$\text{Variance } (t_i - t_m) = \sigma^2 \left[ \frac{1}{A} + \frac{1}{R_1} + \frac{1}{A - (t - q)B} \right. \\ \left. \times \left\{ \frac{B}{A} - \frac{n(R - nb)}{R_1 K} \right\} \right].$$

#### *An example*

The different steps in the analysis of the design have been illustrated by means of the following example. The data analysed have been artificially constructed from a uniformity trial on Malvi cotton at the Institute of Plant Industry, Indore, reported by Hutchinson and Panse (1935), so as to form a progeny row trial with four families having two sister plants in each. Assuming that the seeds available from one of the families and one plant in another family were sufficient for only

twelve replications of each of these progenies, while the remaining plants had each seeds sufficient for sixteen replications, one of the suitable designs for one-way elimination of fertility variation is a super-complete block design with twelve blocks, ten of which accommodate ten plots each, the remaining two having only eight plots each. Blocks have been made unequal for the sake of illustration of the general method of analysis. The eight progenies were allotted at random to the different plots in each of the blocks such that the twenty extra observations from five of the progenies having sixteen replications each, were the second replications within the ten blocks of size ten, and they formed in these blocks a balanced incomplete block design with parameters  $b = 10$ ,  $v = 5$ ,  $r = 4$ ,  $k = 2$  and  $\lambda = 1$ .

The progeny numbers, viz., from 1 to 8, were allotted at random to the different observations taken from twelve consecutive rows of the uniformity trial data. The following table shows the data after rearrangement so as to form a block  $\times$  progeny table:—

*Yield of seed cotton per plot of 1/2000 acre in gm.*

Blocks	Progenies							
	(1) $A_1$	(2) $A_2$	(3) $B_1$	(4) $B_2$	(5) $C_1$	(6) $C_2$	(7) $D_1$	(8) $D_2$
1	82	111	80	93, 80	97, 131	116	95	102
2	69	83	93	63, 50	78	49, 84	85	115
3	129	133	72	95, 133	115	97	89, 90	129
4	70	76	63	107, 76	71	99	107	77, 115
5	120	100	96	69	70, 88	101, 68	63	110
6	74	117	28	83	81, 113	82	91, 86	109
7	86	79	92	103	89, 80	77	72	62, 119
8	94	75	96	90	63	74, 125	76, 119	95
9	130	115	104	77	116	99, 64	45	96, 64
10	63	84	84	40	66	79	51, 111	79, 67
11	85	81	104	76	80	102	42	72
12	111	61	86	73	81	98	71	87



The different parameters in the design are:  $t = 8$ ,  $q = 3$ ,  $n = s = p = 1$ ,  $b' = 2$ ,  $v = 5$ ,  $b = 10$ ,  $r = 4$ ,  $\kappa = 2$ ,  $\lambda = 1$ ,  $R_1 = 12$ ,  $R = 14$ ,  $R_2 = 16$ ,  $K = 10$ ,  $i = 1, 2$ .

The purpose of analysis is first to obtain the error S.S. and the adjusted sum of squares due to progenies, and then to get components of the latter sum of squares to test for family differences.

The different totals required are shown in the table on next page.

$$\Sigma Q_m = -110.45 + 10.95 + 18.35 - 113.25 + 114.05 = -80.35$$

$$A = R_2 - \frac{p^2(r - \lambda)}{K} = 16 - \frac{3}{10} = \frac{157}{10}.$$

$$B = \frac{R(s - n) + p(\lambda p + sr)}{K} = \frac{5}{10} = \frac{1}{2}.$$

$$A - (t - q)B = \frac{157}{10} - \frac{5}{2} = \frac{66}{5}.$$

Hence,

$$t_m = \frac{1}{A} \left\{ Q_m + \frac{B \Sigma Q_m}{A - (t - q)B} \right\} = \frac{10}{157} \left\{ Q_m - \frac{80.35}{26.4} \right\} \\ = \frac{10}{157} (Q_m - 3.0321) \text{ where } m = 4, 5, 6, 7 \text{ and } 8 \text{ and } t$$

stands for the progeny effect.

Thus,

$$t_4 = -7.2282$$

$$t_5 = 0.5043$$

$$t_6 = 0.9757$$

$$t_7 = -7.4065$$

$$t_8 = 7.0712.$$

Again,

$$t_l = \frac{1}{R_1} \left[ Q_l + \frac{n(R - nb) \Sigma Q_m}{K \{A - vB\}} \right] \\ = \frac{1}{12} \left\{ Q_l - \frac{4 \times 80.35}{10 \times \frac{66}{5}} \right\} \\ = \frac{1}{12} (Q_l - 2.4348) \text{ where } l = 1, 2, 3.$$

	1	2	3	4	5	6	7	8	9	10	11	12
Block totals with Nos. of obs.	987 (10)	769 (10)	1082 (10)	861 (10)	885 (10)	864 (10)	859 (10)	907 (10)	910 (10)	724 (10)	642 (8)	668 (8)
Progeny totals with Nos. of obs.	1113 (12)	1115 (12)	998 (12)	1308 (16)	1419 (16)	1414 (16)	1293 (16)	1498 (16)				
Adjusted progeny totals ..	64.45	66.45	-50.55	-110.45	10.95	18.35	-113.25	114.05				

Hence,

$$t_1 = 5.1679.$$

$$t_2 = 5.3346.$$

$$t_3 = -4.4154.$$

The check that  $\Sigma t = 0$  is satisfied.

Now, the adjusted S.S. due to progenies can be obtained from  $\Sigma tQ$  and is equal to 3377.7911.

$$\text{Total S.S. (Cr)} = 943032.$$

$$\text{Block S.S.} = 899308.7000 - 889525.5517 = 9783.1483.$$

$$\text{Error S.S.} = 11402.$$

$$\begin{aligned} \text{Hence interaction S.S.} &= 943032 - 899308.7 - 11402 - 3377.7911 \\ &= 28943.5089. \end{aligned}$$

The sum of squares due to any contrast among the family mean effects of the form  $\Sigma lt = l_1(t_1 + t_2) + l_2(t_3 + t_4) + l_3(t_5 + t_6) + l_4(t_7 + t_8)$  where  $\Sigma l = 0$  can be obtained from  $(\Sigma lt)^2 / \Sigma lt'$ , where  $t_i'$  is a solution of the normal equations obtained by replacing the  $Q$ 's by the  $l$ 's. As the solution of the normal equations is available as function of  $Q$ 's,  $t_i'$  can be easily obtained simply by replacing  $Q_i$  by  $l_i$  in the expression of  $t_i$ . As the variance of  $\Sigma lt$  is  $\sigma^2 \Sigma lt'$ , the divisor for the sum of squares due to the contrast  $\Sigma lt$  must be  $\Sigma lt'$ .

If  $\Sigma mt$  be some other contrast, its sum of squares can be orthogonal to the previous S.S. only if  $\Sigma mt'$  is zero as the covariance between two such contrasts is equal to  $\sigma^2 \Sigma mt'$ .

Hence, all components of the sum of squares, representing family contrasts, can be obtained in this manner.

*Analysis of variance table.*

Due to	d.f.	S.S.	M.S.
Blocks (unadj.) ..	11	9783.15	..
Progenies (adj.) ..	7	3377.79	482.54
Interaction ..	77	28943.51	375.89
Error ..	20	11402.00	570.10

Variances of three typical progeny differences are shown below:—

$$\text{Variance } (t_1 - t_2) = \frac{2}{12} \sigma^2 = .17 \sigma^2.$$

$$\text{Variance } (t_4 - t_5) = \frac{2 \times 10}{157} \sigma^2 = .13 \sigma^2.$$

$$\begin{aligned} \text{Variance } (t_1 - t_4) &= \sigma^2 \left[ \frac{1}{A} + \frac{1}{R_1} + \frac{1}{(A - vB)} \right. \\ &\quad \left. \times \left\{ \frac{B}{A} - \frac{n(R - nb)}{R_1 K} \right\} \right] \\ &= \sigma^2 \left[ \frac{10}{157} + \frac{1}{12} + \frac{5}{66} \left( \frac{5}{157} - \frac{4}{12 \times 10} \right) \right] \\ &= .15 \sigma^2. \end{aligned}$$

In order to test if the first two families differ significantly, we have the contrast  $\Sigma lt$ , where  $l$ 's are 1, 1, -1, -1, 0, 0, 0, 0. The  $t$ 's have been obtained corresponding to these  $l$ 's as

$$t_1' = \frac{8}{99}, t_2' = \frac{8}{99}, t_3' = \frac{-17}{6 \times 33}, t_4' = \frac{-274}{157 \times 26.4},$$

the other  $t$ 's being each equal to  $-10/157 \times 26.4$ ; but it is not necessary to obtain them, as the corresponding  $l$ 's are zero. Evaluation of all the  $t$ 's, however, provides a check as  $\Sigma t' = 0$ .

Thus,  $\Sigma lt' = 4189/15549$  and  $\Sigma lt = 22.1461$ . Now, instead of obtaining the S.S. due to the contrast, the  $t$  test also can be applied to test for the difference between the two families.

Thus,

$$t = \frac{22.1461}{\sqrt{\frac{570 \cdot 10 \times 4189}{15549}}} = 1.78 \text{ on } 20 \text{ d.f.}$$

#### SUMMARY

A generalised balanced design has been defined. The randomised block designs, the balanced incomplete block designs with or without some extra treatments which are present in every block and different balanced super-complete block designs, come out as particular cases of the general design. The design is particularly helpful for single plant progeny row trials with variable amounts of seed available from

the different progenies and for animal experiments involving smaller number of treatments with elimination of litter effect, as in Bio-assays and other animal husbandry experiments. The method of analysis, together with the expressions for finding the standard errors of treatment differences, has been presented. The different steps involved in the analysis of the design have been illustrated by means of an example.

## ACKNOWLEDGEMENT

I feel grateful to Dr. V. G. Panse for some suggestions which helped to improve the quality of the paper.

## REFERENCES

1. Hutchinson, J. B. and Panse, V. G. *Ind. Jour. Agric. Sci.*, 1935, 5, 523-38.
2. ————— .. *Ibid.*, 1937, 7, 531-65.
3. Pearce, S. C. .. *C.A.B. No. 23*, 1953, 32-33.
4. Rao, C. R. .. *J. Amer. Statist. Ass.*, 1947, 42, 541-61.
5. Das, M. N. .. *J. Ind. Soc. Agric. Statist.*, 1953, 5, 161-78.
6. ————— .. *Ibid.*, 1954, 6, 58-76.
7. Giri, N. C. .. *Ibid.*, 1957, 9, 41-51.