DESIGNS BASED ON CONSTANT FREQUENCY DIFFERENCES AND THEIR APPLICATION

By M. N. Das and P. N. Bhargava

Institute of Agricultural Research Statistics, New Delhi

Introduction

One of the authors Das (1958) obtained a method of analysis of a type of non-orthogonal data with constant frequency differences. He indicated that certain designs could be obtained by utilising such property. The application of such designs is more fruitful in those fields where inclusion of more than one block sizes in the same design is not objectionable. As such they are not much suitable for field plot experiment. It may be worth investigating the merits of such designs, where animals are used as experimental units. One such field of research where animals are used as experimental units is bio-assay. A close examination was therefore made to study the utility of such designs for bio-assays.

Efficient designs in bio-assay are somewhat different from those for factorial and varietal trials, because the contrasts of interest are different in the field.

In bio-assay, there are mainly two types of experiments based on quantitative response, one is called parallel line assay and the other slope ratio assay. In the present paper designs suitable for parallel line assay have been considered.

In parallel line assays usually two preparations are taken with say k doses of each preparation. Thus there are $2k$ doses and purpose is to estimate the relative potency of one of the preparations called the test preparation.

Let the $2k$ doses be taken as $2k$-treatments and denoted by t_1, t_2, t_3, ..., t_k for the k doses of the test preparation and s_1, s_2, ..., s_k for the k doses of the standard preparation. Let now L_p denote the contrast $(t_1 + t_2 + t_3 + \ldots + t_k) - (s_1 + s_2 + \ldots + s_k)$ and L_1 the contrast $l_1 (t_1 + s_1) + l_2 (t_2 + s_2) + \ldots + l_k (t_k + s_k)$, where l_i's denote coefficients of linear component among k quantities as taken from the orthogonal polynomial table given in the Fisher-Yates table against k

as the number of observations. It has been shown in the case of symmetrical parallel line assay, *i.e.*, those with the same number of observations, against each of the doses, that the logarithm of the relative potency comes out as a function of L_p/L_1.

Thus the problem for such bio-assay is to obtain some incomplete block designs in which both the contrasts L_p and L_1 can be obtained free from block differences.

The other consideration regarding such assay is to use designs which may provide for more than one block size, though the range of variation of the block sizes should not be wide. Through such designs litters of different sizes can be used in the same experiment and thus some wastage of animal resources avoided. Finney (1952) has also stressed on the need for obtaining such designs.

An attempt has been made in the present paper first to obtain a series of designs based on constant frequency differences and apply them for bio-assay such that both these contrast can be kept free from block differences.

Definition of the Design

Let in a two-way classification with more than one observation per cell, n_{ij} denote the number of observations in the cell defined by the i-th level of one of the factors, say A, and the j-th level of the other factor B, let further $(n_{ij} - n_{mj})$ remain the same, say, X_{im} for the first q levels of B, say, and for the levels $q + 1$ to s let it be, say, Y_{im}, for the remaining levels from $s + 1$ to r it being zero. The number of levels of the factor B evidently varies from 1 to r. A design having this type of frequency for treatments (factor A) and blocks (factor B) may be called designs based on three types of frequency differences. This category of designs can be further generalized if so required by increasing the number of types of differences.

We shall give below the method of analysis for such a design.

Let there be two factors A and B. 'A' having p levels with $a_1, a_2, \ldots, a_p$ effects, B having r levels with effects $b_1, b_2, \ldots, b_q, \ldots, b_s, \ldots, b_r$. The normal equations for estimating a_i after eliminating b_j's come out as

$$Q_i = n_{i.}a_i - \sum_{m=1}^{p} a_m \sum_{j=1}^{r} \frac{n_{ij} \times n_{mj}}{n_{.j}} \quad (i = 1, 2, \ldots, p) \qquad (1)$$

and Q_i is given by

$$Q_i = T_i - \sum \frac{n_{ij} B_j}{n_{.j}}$$

where T_i the i-th treatment total and B_j that of the j-th block.

$$\sum_j n_{ij} = n_i$$

and

$$\sum_i n_{ij} = n_{.j}$$

Eliminating a_p from the first $(p-1)$ equations $i = 1, 2, \ldots, p-1$ with the help of the restriction $\sum_{i=1}^{p} a_i = 0$, these become,

$$n_{i.} a_i - \sum_{m=1}^{p-1} a_m \sum_{j=1}^{r} \frac{n_{ij}(n_{mj} - n_{pj})}{n_{.j}} = Q_i. \tag{2}$$

Let $n_{mj} - n_{pj} = C_m$ for j's varying from 1 to q. $n_{mj} - n_{pj} = C_m'$ for j's varying from $q+1$ to s. and $n_{mj} - n_{pj} = 0$ for j's varying from $s+1$ to r. With these substitutions equation (2) reduces to

$$Q_i = n_{i.} a_i - \sum_{m=1}^{p-1} a_m C_m \left(\sum_{j=1}^{q} \frac{n_{ij}}{n_{.j}} \right) - \sum_{m=1}^{p-1} a_m C_m \left(\sum_{j=q+1}^{s} \frac{n_{ij}}{n_{.j}} \right)$$

$$i = 1, 2, \ldots, p-1.$$

Or

$$Q_i = n_{i.} a_i - R_i \sum_{m=1}^{p-1} a_m C_m - R_i' \sum_{m=1}^{p-1} a_m C_m' \tag{3}$$

where

$$R_i = \left(\sum_{j=1}^{q} \frac{n_{ij}}{n_{.j}} \right)$$

and

$$R_i' = \sum_{\substack{j=q+1 \\ s}}^{s} \frac{n_{ij}}{n_{.j}}.$$

Or

$$a_i - \sum_{m=1}^{p-1} a_m C_m \frac{R_i}{n_{i.}} - \sum_{m=1}^{p-1} a_m C_m' \frac{R_i'}{n_{i.}} = \frac{Q_i}{n_{i.}}.$$

Multiplying the i-th equation by C_i and adding over all such equations, we get,

$$\sum_{i=1}^{p-1} \frac{C_i Q_i}{n_{i.}} = \sum_{m=1}^{p-1} C_m a_m \left(1 - \sum_{i} \frac{R_i C_i}{n_{i.}}\right) - \sum_{m=1}^{p-1} a_m C_m' \sum_{i=1}^{p-1} \frac{R_i' C_i}{n_{i.}} . \tag{5}$$

Similarly multiplying the i-th equation by C_i' and adding over all such equations, we have

$$\sum_{i=1}^{p-1} \frac{C_i' Q_i}{n_{i.}} = -\sum_{m=1}^{p-1} C_m a_m \sum_{i=1}^{p-1} C_i \frac{R_i}{n_{i.}} + \sum_{m=1}^{p-1} a_m C_m' \left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i'}{n_{i.}}\right) \tag{6}$$

from these two equations, *viz.*, (5) and (6), we obtain,

$$\sum_{m=1}^{p-1} a_m C_m = \frac{\sum_{i=1}^{p-1} \frac{C_i Q_i}{n_{i.}} \left(1 - \sum_{i=1}^{p-1} \frac{R_i' C_i}{n_{i.}}\right) + \sum_{i=1}^{p-1} \frac{C_i' Q_i}{n_{i.}} \left(\sum_{i=1}^{p-1} \frac{R_i' C_i}{n_{i.}}\right)}{\left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i}{n_{i.}}\right)\left(1 - \sum_{i=1}^{p-1} \frac{R_i' C_i'}{n_i}\right) - \sum_{i=1}^{p-1} \frac{R_i C_i'}{n_{i.}} \sum_{i=1}^{p-1} \frac{R_i' C_i}{n_{i.}}}$$

$$= L \text{ (say)} \tag{7}$$

$$\sum_{i=1}^{p-1} a_m C_m' = \frac{\sum_{i=1}^{p-1} \frac{C_i Q_i}{n_{i.}} \sum_{i=1}^{p-1} \frac{R_i C_i'}{n_{i.}} + \sum_{i=1}^{p-1} \frac{C_i' Q_i}{n_{i.}} \left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i}{n_{i.}}\right)}{\left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i}{n_{i.}}\right)\left(1 - \sum_{i=1}^{p-1} \frac{R_i' C_i'}{n_{i.}}\right) - \sum_{i=1}^{p-1} \frac{R_i C_i'}{n_{i.}} \sum_{i=1}^{p-1} \frac{R_i' C_i}{n_i}}$$

$$= M \text{ (say)} \tag{8}$$

Substituting these values, *viz.*, (7) and (8) in (4), we have,

$$a_i = \frac{Q_i}{n_{i.}} + \frac{R_i}{n_{i.}} L + \frac{R_i'}{n_{i.}} M. \tag{9}$$

Solution to a_p can be obtained from $\Sigma a_i = 0$.

Variance of the $(a_i - a_m)$ can be obtained by finding the coefficients of Q_i and Q_m in the estimate of $(a_i - a_m)$ where $m \leqslant (p-1)$.

When $(n_{ij} - n_{mj})$ takes only two values for different j's, *viz.*, X_{im} and 0, *i.e.*, $C_m = C_m'$ and this reduces to the case discussed by Das (1958). The equation (3) reduces in this case to the form,

$$n_{i.}a_i - \sum_{m=1}^{p-1} a_m C_m R_i = Q_i \tag{10}$$

where

$$R_i = \sum_{j=1}^{s} \frac{n_{ij}}{n_{.j}}.$$

Solution to a_i, *viz.*, (9) reduces to

$$a_i = \frac{Q_i}{n_{i.}} + \frac{R_i}{n_{i.}} \chi \tag{11}$$

where

$$\chi = \frac{\sum_{i=1}^{p-1} \frac{C_i Q_i}{n_{i.}}}{\left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i}{n_{i.}}\right)}$$

and the variance of $(a_i - a_m)$ is

$$V(a_i - a_m) = \sigma^2 \left\{\frac{1}{n_{i.}} + \frac{1}{n_{m.}}\right\} + \frac{\left[\left(\frac{R_i}{n_{i.}} - \frac{R_m}{n_{m.}}\right)\left(\frac{C_i}{n_{i.}} - \frac{C_m}{n_{m.}}\right)\right]}{\left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i}{n_{i.}}\right)} \tag{12}$$

and the corresponding,

$$V(a_i - a_p) = \sigma^2 \left\{ \frac{Q}{n_{i.}} + \frac{\frac{C_i}{n_{i.}}\left(\frac{R_i}{n_{i.}} - \sum_{i=1}^{p-1} \frac{R_i}{n_{i.}}\right)}{\left(1 - \sum_{i=1}^{p-1} \frac{R_i C_i}{n_{i.}}\right)} \right\} \qquad (13)$$

By using this property of non-orthogonal data, we can obtain several series of designs, which are particularly suitable for parallel line assays. A list of such designs is given below. In these designs both the preparation as also the regression contrasts are unaffected by block effects and the other components required for validity tests can be obtained through the method given by Das (1958). We have given below as illustration some of the designs. In these designs only the block types have been shown. Some of these block types may be omitted if necessary, or some of the types of blocks may be repeated any number of times, if so required. Through each of these designs the preparation and regression contrasts can be estimated from the corresponding contrasts of the dose means.

(i) *Designs for the four-point parallel line assays.*—The figures in the body of the table denote frequencies of observations, s_1, s_2 denote the two doses of the standard preparation, and t_1 and t_2 denote those of the test preparation.

Types of Blocks	Dose				Size of the block
	s_1	s_2	t_1	t_2	
	0	1	1	0	2
	1	2	2	1	6
	2	3	3	2	10
	.	.	.	.	.
	1	1	1	1	4
	2	2	2	2	8
	.	.	.	.	.
Totals of observation	S_1	S_2	T_1	T_2	

In this design five blocks having only two types of differences have been presented. Depending on the availability of animals, we may repeat any type of blocks any number of times or omit one type altogether. Blocks of sizes two only presented above cannot themselves form designs, though any n different blocks where $n \geqslant 1$ can constitute a design. Some further blocks may also be obtained, if necessary by increasing the frequencies in any types of block by a constant. If S_1, S_2 and T_1, T_2 denote the observation totals against the four doses, it is easily seen that the preparation contrast of dose effects $L_p = (s_1 + s_2) - (t_1 + t_2) =$ the corresponding contrasts of means, independent of block differences. Further, L_1 (the regression contrast) $= -(s_1 + t_1) + (s_2 + t_2)$ is also independent of block differences. Thus, through such a design the potency can be estimated without any loss of information.

(ii) It will be seen that in the previous design, we have used two sets of initial blocks, one incomplete and the other complete. The other blocks have been developed from them by adding a frequency unity or any other suitable number to each treatment. Thus these designs are based on two types of frequency differences.

The usefulness of the design can be increased further, by making block sizes closer. This can be achieved if one more type of incomplete blocks can be accommodated in the same design. It will then be a design based on three types of frequency differences, for the purpose of illustration the initial blocks for a six-point assay have been taken and described below.

(*a*) *Design for a six-point assay*

Types of blocks	Doses						Block size
	s_1	s_2	s_3	t_1	t_2	t_3	
1	0	0	1	1	0	0	$2 + 6i$
2	1	0	1	1	0	1	$4 + 6j$
3	1	1	1	1	1	1	$6 + 6k$

Here (1), (2) and (3) are the initial blocks from which the other blocks can be generated as stated by adding unity to each frequency. The method of analysis of the data obtained through such designs

follows the same line as indicated for (1). In this case $L_p = (s_1 + s_2 + s_3) - (t_1 + t_2 + t_3)$ and $L_1 = -(s_1 + t_1) + (s_3 + t_3)$; it will be clear that they are free from block differences.

(*b*) *Design for eight-point assays*

Similarly for an eight-point assay we can use the design

Types of blocks	Doses								Possible block sizes
	s_1	s_2	s_3	s_4	t_1	t_2	t_3	t_4	
1	0	0	0	1	1	0	0	0	$2 + 8i$
2	1	0	0	1	1	0	0	1	$4 + 8j$
3	1	1	1	1	1	1	1	1	$8 + 8k$
Total of observations	S_1	S_2	S_3	S_4	T_1	T_2	T_3	T_4	

where i, j and k can take the values 0, 1, 2. ... so as to generate different blocks.

SUMMARY

If in a two-way classification with more than one observation per cell, n_{ij} denote the number of observations in the cell defined by the i-th class of A and j-th class of the other factor B, then it is well known that the data are orthogonal and can be analysed orthogonally if n_{ij}/n_{mj} remains the same for all values of j, i and m denoting any two levels of the factor A. It has been shown in this paper that if $(n_{ij} - n_{mj})$ remains the same, for some j's and takes the value C_{im} and for some other it takes the values C_{im}' and is zero for the other j's corresponding to a given section of the levels of B, then the data though non-orthogonal can be analysed easily since this type of data can be so manipulated as to make the normal equations solution easy and systematic, required for the analysis of variance of such data.

For the case when C_{im} and C_{im}' are equal, the algebraic solution of the normal equation is easy, and has also been discussed.

Utilising the above property we have presented two designs for three types of frequency differences suitable for four and six point assay.

REFERENCES

1. Das, M. N. .. *A New Type of Design*, 11th Annual General Meeting of the Indian Society of Agricultural Statistics, held at Bangalore in January, 1958.

2. Finney, D. J. .. *Statistical Method in Biological Assay*, 1952.