ANALYSIS OF A BALANCED INCOMPLETE TWO-WAY DESIGN

A. M. KSHIRSAGAR & P. S. SIMHA

(Received Sept. 13, 1967)

Abstract

A design in which heterogeneity is eliminated in two directions is said to be balanced incomplete if both rows as well as columns are incomplete and if all elementary treatment contrasts are estimated with the same precision. Kshirsagar [1] and Pearce [3] have given illustrations of such designs. In this paper, the complete analysis—with recovery of inter-row and inter-column information—of such a design is given along with the estimates of the inter-row and inter-column variances and is illustrated by a numerical example. Kshirsagar's condition of balancing in two-way designs is in terms of latent roots. For designs with a particular structures this condition can be expressed in terms of the parameters of the design. This is also done in this paper.

1. Introduction

Let there be uu' plots arranged in u rows and u' columns and let v treatments be assigned to these plots in such a way that every treatment is replicated r times and the ith treatment occurs l_{ij} times ($l_{ij}=0$ or 1 only) in the jth row and m_{ik} times ($m_{ik}=0$ or 1) in the kth column ($j=1,2,\cdots,u$; $k=1,2,\cdots,u'$). The $v\times u$ and $v\times u'$ matrices of l_{ij}'s and m_{ik}'s will be denoted by $\boldsymbol{L}$ and $\boldsymbol{M}$ respectively. They are called the row and column incidence matrices respectively. The set-up is:

$$y_{ij}=\mu+\alpha_i+\beta_j+\tau_s+e_{ij} \tag{1.1}$$

where

y_{ij}=yield of the plot in the ith row and jth column, (which receives sth treatment)

μ=general mean,

α_i=effect of ith row,

β_j=effect of jth column,

τ_s=effect of sth treatment,
e_{ij}=error.

In the intra-row and column analysis, α_i, β_j are regarded fixed and e_{ij}'s are normally and independently distributed with zero means and variance σ^2. If inter-row and column information are also to be recovered, α_i, β_j are assumed to be independently and normally distributed with zero means and variances σ_r^2 and σ_c^2 respectively; all α_i, β_j and e_{ij} are independent. The 'intra' estimate of a contrast $\tau_s-\tau_k$ will be denoted by $\hat{t}_s-\hat{t}_k$ where $\hat{t}_s$ are any solution of the equations

$$Q=F\hat{t} \tag{1.2}$$

where

$\boldsymbol{Q}$=the column vector of Q_i's
$\boldsymbol{t}$=the column vector of t_i's
$\boldsymbol{R}$=the column vector of R_i's
$\boldsymbol{C}$=the column vector of C_j's
$\boldsymbol{T}$=the column vector of T_s's
$R_i=\sum_{j=1}^{u'} y_{ij}$, $C_j=\sum_{i=1}^{u} y_{ij}$, T_s=total of sth treatment
$g=\sum_i \sum_j y_{ij}$, $\boldsymbol{E}_{pq}$=any $p\times q$ matrix with all unit elements.

$$\boldsymbol{Q}=\boldsymbol{T}-\frac{1}{u'}\boldsymbol{LR}-\frac{1}{u}\boldsymbol{MC}+\frac{g}{v}\boldsymbol{E}_{v_1} \tag{1.3}$$

$$\boldsymbol{F}=r\boldsymbol{I}-\frac{1}{u'}\boldsymbol{LL}'-\frac{1}{u}\boldsymbol{MM}'+\frac{r}{v}\boldsymbol{E}_{vv}\,. \tag{1.4}$$

The combined intra and inter-row and column estimates of contrasts $\tau_s-\tau_k$ are t_s-t_k where t_s's are any solutions of

$$\boldsymbol{P}=\left(W\boldsymbol{F}+\frac{1}{u'}W_r\boldsymbol{LL}'+\frac{1}{u}W_c\boldsymbol{MM}'\right)\boldsymbol{t} \tag{1.5}$$

$$=\boldsymbol{Kt}\,,\quad \text{say}$$

where

$$\boldsymbol{P}=W\boldsymbol{Q}+W_r\boldsymbol{Q}_r+W_c\boldsymbol{Q}_c\,, \tag{1.6}$$

$$\boldsymbol{Q}_r=\frac{1}{u'}\boldsymbol{LR}-\frac{g}{v}\boldsymbol{E}_{v_1} \tag{1.7}$$

$$\boldsymbol{Q}_c=\frac{1}{u}\boldsymbol{MC}-\frac{g}{v}\boldsymbol{E}_{v_1} \tag{1.8}$$

$$W=1/\sigma^2,\quad W_r=1/(\sigma^2+u'\sigma_r^2)\quad \text{and}\quad W_c=1/(\sigma^2+u\sigma_c^2)\,. \tag{1.9}$$

2. Analysis of a balanced design

Kshirsagar [1] has given a two-way design in which the rows form a Partially Balanced Incomplete Block (PBIB) design with 2 associate classes and the columns also form a PBIB design with 2 associate classes and the design is balanced on the whole. The lay-out of the design along with fictitious yields is given below: (Figures in brackets indicate treatment numbers.)

(2.0)

16.077 (2)	15.528 (4)	18.572 (8)	9.589 (7)	12.364 (6)	18.591 (3)
21.387 (3)	13.930 (5)	18.496 (7)	15.761 (2)	18.260 (4)	33.365 (9)
13.262 (5)	9.619 (6)	14.673 (3)	11.786 (1)	10.857 (7)	17.998 (8)
17.898 (4)	22.213 (9)	15.811 (1)	18.904 (8)	16.157 (3)	16.238 (5)
14.181 (6)	8.536 (7)	27.624 (9)	9.997 (5)	9.334 (1)	18.279 (2)
15.825 (1)	17.788 (8)	17.028 (2)	14.411 (4)	25.326 (9)	16.935 (6)

One can easily write down the matrices L and M from the plan of the design. From these, we readily obtain

(2.1) $$LL' = 3I_1 \times I_2 - J_1 \times I_2 - I_1 \times J_2 + 3J_1 \times J_2$$

and

(2.2) $$MM' = J_1 \times I_2 + I_1 \times J_2 + 2J_1 \times J_2$$

where

$$I_i = \text{identity matrix of order } m_i$$
$$J_i = E_{m_i m_i} \qquad (i=1,2)$$
$$v = m_1 m_2$$

and $\times$ stands for the Kronecker product of two matrices. For this design, $v=9$, $r=4$, $u=u'=6$, $m_1=m_2=3$. Kshirsagar [1] has only indicated the method of analysis of such designs but has not given explicit expressions for t and $\hat{t}$, nor has he given the estimates of σ_r^2 and σ_c^2.

Substituting from (2.1) and (2.2) in (1.2) we find that

$$F=\frac{7}{2}\left(I-\frac{1}{9}E_{vv}\right) \tag{2.3}$$

and hence, with the usual restriction $E_{1v}\hat{t}=0$, a solution of (1.2) is

$$\hat{t}=\frac{2}{7}Q\,. \tag{2.4}$$

These provide the 'intra' estimates of the τ_s's. We now define

$$K^*=a_1I_1\times I_2+a_2I_1\times J_2+a_3J_1\times I_2+a_4J_1\times J_2 \tag{2.5}$$

and determine the arbitrary constants a_i in such a way that K^* is the pseudo-inverse of K (when (2.1) and (2.2) are substituted in (1.5)) i.e.

$$K^*K=I-\frac{1}{v}E_{vv} \quad \text{i.e.} \quad I_1\times I_2-\frac{1}{v}J_1\times J_2\,.$$

A little algebra shows that

$$a_1=\frac{2}{7W+W_r}\,, \qquad a_2=a_3=\frac{2(W_r-W_c)}{3(7W+W_c)(7W+W_r)}\,. \tag{2.6}$$

Consequently, premultiplying (1.5) by K^* and using $E_{1v}t=0$, we obtain (as $E_{1v}P=0$)

$$t=[a_1I_1\times I_2+a_2(I_1\times J_2+J_1\times I_2)]P\,. \tag{2.7}$$

This procedure is analogous to the one employed by Zelen and Federer [5] in analysis of designs possessing property (A).

A two-way design is said to possess property (A), if

$$LL'=\sum_\delta h_r(\delta)D^\delta\,; \quad MM'=\sum_\delta h_c(\delta)D^\delta \tag{2.8}$$

where

$$\delta=(\delta_1,\delta_2,\cdots,\delta_n)\,, \qquad \delta_i=0 \quad \text{or} \quad 1$$

$$D^\delta=D_1^{\delta_1}\times D_2^{\delta_2}\times\cdots\times D_n^{\delta_n}$$

and

$$D_i^{\delta_i}=\begin{cases} I_i\,; & \delta_i=0\,,\\ J_i\,; & \delta_i=1\,.\end{cases}$$

It is well-known that the variance-covariance matrix of Q is σ^2F and that of P is

$$WF+W_r\left(\frac{1}{u'}LL'-\frac{r}{v}E_{vv}\right)+W_c\left(\frac{1}{u}MM'-\frac{r}{v}E_{vv}\right). \tag{2.9}$$

From this, the variances of all elementary contrasts $\hat{t}_s-\hat{t}_k$ or of t_s-t_k (i.e. with recovery of inter-row and column information) can be easily seen to be as below:

(I) Without recovery: The variance of any elementary contrast is $4\sigma^2/7$.

(II) If $s=(s_1, s_2)$, $k=(k_1, k_2)$ and treatments 1 to 9 are renumbered as $(1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $\cdots$, $(3,3)$ respectively,

$$V(t_s-t_k)=\frac{4}{3}\left[\frac{1}{7W+W_r}+\frac{1}{7W+W_c}+\frac{2}{7W+\phi W_r+(1-\phi)W_c}\right] \tag{2.10}$$

where

$$\phi=\begin{cases}1 & \text{if } s_1=k_1 \text{ or } s_2=k_2,\\ 0 & \text{otherwise.}\end{cases}$$

3. Estimation of σ^2, σ_r^2 and σ_c^2

It is well known that σ^2 can be estimated from the error mean square in the 'intra' analysis i.e. by

$$\hat{\sigma}^2=\frac{\sum_i\sum_j y_{ij}^2-\frac{1}{u'}R'R-\frac{1}{u}C'C+\frac{g^2}{uu'}-Q'\hat{t}}{uu'-u-u'-v}. \tag{3.1}$$

Estimates of σ_r^2 and σ_c^2 can be found from S_r, the sum of squares due to rows (adjusted for columns and treatments) and S_c, the sum of squares due to columns (adjusted for rows and treatments). Roy and Shah [2] have shown that

$$S_r=\frac{R'R}{u'}-\frac{g^2}{uu'}+Q'\hat{t}-\left(T-\frac{1}{u}MC\right)'\left(rI_v-\frac{1}{u}MM'\right)^*\left(T-\frac{1}{u}MC\right) \tag{3.2}$$

where $*$ denotes the pseudo-inverse of a matrix. This pseudo-inverse can be found out in the same manner as in the earlier section. After a little algebra, we find

$$E(S_r)=(u-1)(\sigma^2+u'\sigma_r^2)-\sigma_r^2 t_r LL'\left(rI_v-\frac{1}{u}MM'\right)^* \tag{3.3}$$

$$=5\sigma^2+27\sigma_r^2 \qquad \text{(for the design (2.0)).}$$

In the same manner, we find

$$E(S_c)=5\sigma^2+27\sigma_c^2. \tag{3.4}$$

Estimates of W, W_r and W_c can, therefore, be obtained from (3.1), (3.3) and (3.4). If it so happens that the estimates of σ_r^2 or σ_c^2 turn out to be negative, alternative estimates suggested by Roy and Shah [2] may be used.

4. Numerical example

For the yields, given in (2.0), we obtain the following values of $\boldsymbol{t}$ and $\hat{\boldsymbol{t}}$. They have been obtained by computing the required quantities such as $\boldsymbol{Q}$, $\boldsymbol{P}$, error S.S., S_r and S_c and using (2.4) and (2.7),

$$\hat{\boldsymbol{t}}'=[-2.507,\ -1.045,\ -0.093,\ 0.295,\ -3.192,\ -2.403,\ -3.367,\ 2.358,\ 9.954]$$

$$\boldsymbol{t}'=[-2.202,\ 0.577,\ -0.800,\ 2.971,\ -2.643,\ -2.059,\ -3.545,\ 1.890,\ 8.431].$$

It is to be borne in mind that only contrasts of the treatment effects are estimable and the best estimates are obtained by substituting the above values in the contrasts. The analysis of variance table is as follows.

Analysis of variance table

Source	d.f.	s.s.	M.S.	F
Rows (unadjusted)	5	208.772	41.754	
Columns (unadjusted)	5	199.154	39.831	
Treatments (adjusted for rows & columns)	8	487.899	60.988	30.46
Error (by subtraction)	17	34.032	2.002	
Total	35	929.857		

The row s.s. (adjusted for columns and treatments) is 90.927. The column s.s. (adjusted for rows and treatments) is 15.024. Hence the estimates of σ^2, σ_r^2 and σ_c^2 are respectively 2.266, 2.948 and 0.137. The estimated variance of any elementary treatment contrast without recovery of inter-row and column information is 0.572. If inter-row and column information is recovered, the estimated variance of $t_{(ij)}-t_{(i'j')}$ is 0.931 if either $i=i'$ or $j=j'$ and 0.905 if $i\neq i'$ and $j\neq j'$.

5. Pearce's balanced design

Kshirsagar [1] had remarked that the design (2.0) is the only design known to him, in which both rows and columns are incomplete and has an over-all balance. However, later Pearce [3] came forward with one more such design and Preece [4] has shown that the two designs—the one given by Pearce and (2.0)—differ structurally. It can, therefore be seen that for Pearce's design

$$\boldsymbol{LL}' = \boldsymbol{I}_1 \times \boldsymbol{J}_2 + \boldsymbol{J}_1 \times \boldsymbol{I}_2 + 2\boldsymbol{J}_1 \times \boldsymbol{J}_2 ,$$

$$\boldsymbol{MM}' = 3\boldsymbol{I}_1 \times \boldsymbol{I}_2 - \boldsymbol{I}_1 \times \boldsymbol{J}_2 - \boldsymbol{J}_1 \times \boldsymbol{I}_2 + 3\boldsymbol{J}_1 \times \boldsymbol{J}_2 .$$

i.e. they are the same as (2.1) and (2.2) with $\boldsymbol{L}$ and $\boldsymbol{M}$ interchanged. Hence the analysis of Pearce's design will be the same as that of (2.0); only rows and columns have to be interchanged.

6. Balancing in two-way designs

Kshirsagar [1] has obtained a necessary and sufficient condition for balancing in two-way designs. His condition is in terms of the latent roots of certain matrices associated with the rows and columns. In this section, we obtain an alternative condition for balancing. This condition is valid only for designs which possess property (A); but its advantage is that it is in terms of the parameters of the design rather than the latent roots.

For designs possessing property (A),

$$\boldsymbol{F} = \sum_{\delta} g(\delta) \boldsymbol{D}^{\delta} \tag{6.1}$$

where $\delta = (\delta_1, \cdots, \delta_n)$, each $\delta_i = 0$ or 1, the summation is over all binary numbers δ and $g(\delta)$ are certain constants.

From (2.8) and (6.1),

$$g(0, 0, \cdots, 0) = r - h_r(0, \cdots, 0)/u' - h_c(0, \cdots, 0)/u = g(0) \qquad \text{(say)},$$

$$g(1, 1, \cdots, 1) = r/v - h_r(1, \cdots, 1)/u' - h_c(1, \cdots, 1)/u = g(1) \qquad \text{(say)}$$

and

$$g(\delta) = -h_r(\delta)/u' - h_c(\delta)/u \tag{6.2}$$

for any other δ.

For a balanced design $\boldsymbol{F}$ must be of the form (see for example Kshirsagar [1])

$$\lambda\left(I_v - \frac{1}{v}E_{vv}\right) \tag{6.3}$$

where λ is the same constant.

This will be so if and only if

$$\frac{h_r(\delta)}{u'} + \frac{h_c(\delta)}{u} = 0 \tag{6.4}$$

for all δ, excluding $\delta=(0,\cdots,0)$ and $\delta=(1,1,\cdots,1)$. This is, therefore, a necessary and sufficient condition for balancing. That the condition is necessary can readily be seen by comparing (6.1) and (6.3). Conversely, if (6.4) is satisfied, from (6.1), we obtain

$$F = g(0)I_v + g(1)E_{vv}$$

but $FE_{v_1}=0$ and hence

$$g(0) + vg(1) = 0 .$$

This, therefore, on substitution in F, yields

$$F = g(0)\left[I - \frac{1}{v}E_{vv}\right].$$

This, therefore, proves that the condition (6.4) is both necessary and sufficient.

Acknowledgements

The authors are grateful to Dr. N. K. Chakravarti, Asst. Director, Defence Science Laboratory, for encouragement and to the Director, Defence Science Laboratory, Delhi for permission to publish this paper.

DEFENCE SCIENCE LABORATORY, INDIA

REFERENCES

[1] A. M. Kshirsagar, "On balancing in designs in which heterogeneity is eliminated in two directions," *Bull. Calcutta Statist. Ass.*, 7 (1957), 161-166.

[2] J. Roy and K. R. Shah, "Analysis of two-way designs," *Sankhyā A*, 23 (1961), 129-145.

[3] S. C. Pearce, "The use and classification of non-orthogonal designs," *J. R. Statist. Soc. A*, 126 (1963), 353-377.

[4] D. A. Preece, "Classifying Youden rectangles," *J. R. Statist. Soc. B*, 28 (1966), 118-130.

[5] M. Zelen and W. Federer, "Application of the calculus of factorial arrangements, II," *Ann. Math. Statist.*, 35 (1964), 658-672.