MISCELLANEOUS

ON THE EFFICIENCY FACTOR OF BLOCK DESIGNS

By J. ROY

Indian Statistical Institute, Calcutta

1. Introduction and summary

Suppose that a comparative trial involving v treatments has to be carried out using n experimental units arranged in b blocks. To choose one out of the many possible designs for such an experiment, the criterion generally adopted is the 'efficiency factor' of the design. This is defined as a ratio $E = V_R/\bar{V}$ where $\bar{V}$ is the average variance of the intra-block esimate of the difference between the effects of any two treatments for the design under consideration and V_R that for a Randomised Block design using the same number of experimental units, V_R and $\bar{V}$ being computed on the assumption that the intra-block error variance is the same in both the cases. The problem that arises is this : which one of the possible designs for an experiment with v treatments with n experimental units arranged in b blocks has the highest efficiency factor ?

Kempthorne (1956) derived an explicit expression for the efficiency factor of a special type of designs—to be called 'connected proper equireplicate binary design' in this paper. A more general expression is deduced here which leads to Kempthorne's result as a special case. An upper bound for the efficiency factor is derived. A design which attains this upper bound is defined to be 'most efficient'.

It is shown that in a most efficient design the difference between the effects of any two treatments is estimable with the same precision, and vice-versa. As a corollary, it follows that in the class of proper binary designs, a most efficient design, if it exists, is necessarily a Balanced Incomplete Block design.

New designs are sometimes constructed from given proper equireplicate binary designs by dualising, that is by interchanging the role of blocks and treatments. A simple expression is derived connecting the efficiency factor of the dual design with that of the original design, from which an interesting result follows, namely that the efficiency factor of the dual design is greater than, equal to, or less than the efficiency factor of the original design according as the number of blocks in the original design is greater than, equal to, or less than the number of treatments in the original design.

2. A theorem on linear estimation

It is well-known (Rao, 1952) that in problems of linear estimation of parameters involved linearly in the expectations of uncorrelated random variables with the same variance σ^2, the method of least squares leads to the algebraically consistent system of normal equations

$$\theta C = Q \qquad \ldots \quad (2.1)$$

where $\theta(: 1\times m)$ are unknown parameters and $Q(: 1\times m)$ are known linear functions of the given random variables. The dispersion matrix of Q is $\sigma^2 C$.

If C is non-singular, the solution of (2.1) may be written as

$$\hat{\theta} = QD \qquad \ldots \quad (2.2)$$

with the dispersion matrix of $\hat{\theta}$ given by

$$D(\hat{\theta}) = \sigma^2 . D \qquad \ldots \quad (2.3)$$

where $D = C^{-1}$. If C is singular, of rank c, $c < m$ say, the results (2.2) and (2.3) still remain valid provided D is interpreted as a pseudo-inverse of C (Rao, 1955). A particular pseudo-inverse is

$$D = \Gamma \nabla \Gamma' \qquad \text{... (2.4)}$$

where Γ is an orthogonal matrix and ∇ a diagonal matrix with diagonal elements $(\delta_1^{-1}, \delta_2^{-1}, ..., \delta_c^{-1}, 0, 0, ..., 0)$ where δ_t $(t = 1, 2, ..., c)$ are the positive latent roots of C. This is the case where the normal equations (2.1) are solved by specifying algebraically consistent non-stochastic values of $(m-c)$ linearly independent functions of the parameters. We thus get the following :

Theorem 2.1: *The positive latent roots of the dispersion matrix of a solution of the normal equations are given by* σ^2/δ_t *where* δ_t $(t = 1, 2, ..., c)$ *are the positive latent roots of the matrix* C.

3. Some definitions

By a 'design' we shall understand an allocation of v treatments, one on each of n experimental units (eu's) arranged in b blocks or groups of eu's. The matrix of the form $v \times b$,

$$N = ((n_{ji})) \qquad \text{... (3.1)}$$

where n_{ji} denotes the number of eu's in the i-th block getting the j-th treatment will be called the 'incidence-matrix' of the design. The number of replications of the j-th treatment is given by

$$r_j = \sum_{i=1}^{b} n_{ji} \qquad \text{... (3.2)}$$

and the number of eu's in the i-th block by

$$k_i = \sum_{j=1}^{v} n_{ji}. \qquad \text{... (3.3)}$$

Obviously,

$$n = \sum_{j=1}^{v} r_j = \sum_{i=1}^{b} k_i. \qquad \text{... (3.4)}$$

we shall write

$$R = \text{diag } (r_1, r_2, ..., r_v) \qquad \text{... (3.5)}$$

$$K = \text{diag } (k_1, k_2, ..., k_b) \qquad \text{... (3.6)}$$

and call

$$C = R - NK^{-1}N' \qquad \text{... (3.7)}$$

the 'coefficient matrix' of the design.

Since each row (or column) of C adds upto zero, the rank of C is at most $(v-1)$. A design will be said to be 'connected' if the rank of its coefficient matrix is $(v-1)$.

A design will be said to be a 'binary' design if a treatment occurs at most once in a block that is if n_{ij} takes one of the only two possible values 1 or 0. For a binary design, the matrix

$$\Lambda = ((\lambda_{jj'})) \qquad \text{... (3.8)}$$

where $\lambda_{jj'}$ denotes the number of blocks in which treatments j and j' occur together ($j \neq j' = 1, 2, ..., v$) and $\lambda_{jj} = r_j$ will be said to be the 'association-matrix' of the design. Obviously

$$\Lambda = NN'. \quad \text{... (3.9)}$$

A binary design will be called 'singular' if its association-matrix is singular.

A design in which $k_1 = k_2 = ... k_b$ holds, will be said to be a 'proper' design, the common value of the k_i's being denoted by k. A design in which $r_1 = r_2 = ... = r_v$ holds will be said to be an 'equireplicate' design and the common value of the r_j's will be denoted by r. A design with $k_i < v$ for $i = 1, 2, ..., b$ will be called an 'incomplete block' design. A proper equireplicate design will be said to be 'symmetric' if $b = v$.

A proper equireplicate incomplete block design in which $\lambda_{jj'} = \lambda$ for all $j \neq j'$ is said to be a Balanced Incomplete Block (BIB) design.

A design D^* will be said to be the 'dual' of a design D if the incidence matrix N^* of the design D^* is the transpose of the incidence matrix N of the design D, that is, if $N^* = N'$. Obviously between a design and its dual, the roles of blocks and treatments are merely interchanged.

4. The efficiency factor of a block design

Suppose in an experiment with a given connected design, the total yield of the j-th treatment is T_j and that of the i-th block is B_i. Write

$$T = (T_1, T_2, ..., T_v)$$

$$B = (B_1, B_2, ..., B_b)$$

and define the 'adjusted yields' by

$$Q = T - BK^{-1}N'$$

where K and N are defined in (3.5) and (3.1). Denote the effect of the j-th treatment by θ_j and write

$$\theta = (\theta_1, \theta_2, ..., \theta_v).$$

Then under the assumption that block and treatment effects are additive, the 'intra-block' equations for estimating the treatment effects are given by

$$\theta C = Q \quad \text{... (4.1)}$$

where C is the coefficient matrix of the design defined in (3.7). The dispersion matrix of Q is known to be $C.\ \sigma^2$ where σ^2 is the intra-block error.

Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_v)$ denote any particular solution of (4.1). It follows from section 2 that the solution must satisfy

$$\hat{\theta}_1 + \hat{\theta}_2 + ... + \hat{\theta}_v = a \quad \text{... (4.2)}$$

where a is some pre-assigned constant. Let $D\,.\,\sigma^2 = ((d_{jj'}))\,.\,\sigma^2$ denote the dispersion-matrix of $\hat{\theta}$. It follows from (4.2) that

$$\sum_{j'=1}^{v} d_{jj'} = 0 \text{ for } j = 1, 2, ... v. \quad \text{... (4.3)}$$

The 'efficiency factor' E of a connected design is defined by

$$E = \frac{2\sigma^2}{(n/v)} / \bar{V} \quad \dots \quad (4.4)$$

where

$$\bar{V} = \frac{1}{v(v-1)} \sum_{j \neq j'=1} \sum \text{var}\, (\hat{\theta}_j - \hat{\theta}_{j'}). \quad \dots \quad (4.5)$$

Obviously, therefore,

$$v(v-1) \quad \bar{V} = \sum_{j=1}^{v} \sum_{j'=1}^{v} (d_{jj} + d_{j'j'} - 2d_{jj'})\sigma^2$$

$$= 2v(\sum_{j=1}^{v} d_{jj})\sigma^2 \qquad \text{because of (4.3).}$$

Thus

$$\bar{V} = \frac{2\sigma^2}{v-1} \text{ tr } D \quad \dots \quad (4.6)$$

where 'tr' denotes the trace of a matrix. Now the trace of a matrix is equal to the sum of its latent roots. Using Theorem 2.1 it follows that

$$\bar{V} = \frac{2\sigma^2}{v-1} \sum_{t=1}^{v-1} \frac{1}{\delta_t} \quad \dots \quad (4.7)$$

where $\delta_t (t = 1, 2, \dots, v-1)$ are the positive latent roots of the matrix C. Substituting in (4.4) we obtain the following.

Theorem 4.1: *The efficiency factor E of a connected design is given by*

$$E = \frac{v}{n} \bar{\delta} \quad \dots \quad (4.8)$$

where $\bar{\delta}$ is the harmonic mean of the positive latent roots of the coefficient matrix of the design.

For a proper equireplicate binary design the coefficient-matrix is given by

$$C = r\, I - \frac{1}{k} \Lambda \quad \dots \quad (4.9)$$

where I is the identity matrix and Λ the association matrix of the design. The latent roots δ_t of C are connected with the latent roots ϕ_t of Λ by the identity

$$\delta_t = r - \frac{\phi_t}{k} \; (t = 1, 2, \dots, v). \quad \dots \quad (4.10)$$

The dominant latent root of Λ is rk and will be denoted by $\phi = rk$. If the design is connected all other latent roots of Λ are smaller than rk. Making use of Theorem 4.1 and the formula (4.10) we get the following :

Corollary 4.1: (Kempthorne, 1956) : *The efficiency factor E of a connected proper equireplicate binary design is given by*

$$E = \frac{(v-1)}{rk} \Big/ \sum_{t=1}^{v-1} \frac{1}{rk - \phi_t} \quad \dots \quad (4.11)$$

where $\phi_t (t = 1, 2, \ldots, v-1)$ are the latent roots other than rk of the association matrix of the design.

If the design is singular and the association matrix has zero latent root of multiplicity u, the efficiency factor can be written as

$$E = \frac{(v-1)}{rk} \left\{ \frac{u}{rk} + \sum_{t=1}^{v-u-1} \frac{1}{rk-\phi_t} \right\} \qquad \ldots \quad (4.12)$$

where $\phi_t (t = 1, 2, \ldots, v-u-1)$ are the latent roots other than 0 and rk of the association matrix.

5. Characterization of most efficient designs

Since the harmonic mean of a set of positive quantities is never greater than the arithmetic mean, it follows that

$$\delta \leqslant \frac{1}{v-1} \sum_{i=1}^{v-1} \delta_i = \frac{1}{v-1} \, t_r C. \qquad \ldots \quad (5.1)$$

But,

$$t_r C = \sum_{j=1}^{v} \left\{ r_j - \sum_{i=1}^{b} \frac{n_{ji}^2}{k_i} \right\} = n - \sum_{i=1}^{b} \frac{m_i}{k_i} \qquad \ldots \quad (5.2)$$

where

$$m_i = \sum_{j=1}^{v} n_{ji}^2 \geqslant \max\left(k_i, \frac{k_i^2}{v} \right) \qquad \ldots \quad (5.3)$$

Let b' be the number of blocks each having less than v cu's and n' the total number of cu's in blocks each having v or more cu's. Then

$$t_r C \leqslant n - b' - \frac{n'}{v} \leqslant n - b. \qquad \ldots \quad (5.4)$$

Summarizing, we get the following:

Theorem 5.1: *The efficiency factor E of any connected design cannot exceed E_0 given by*

$$E_0 = \frac{v(n-b')-n'}{n(v-1)} \qquad \ldots \quad (5.5)$$

For an incomplete block design $k_i < v$ and therefore $m_i > k_i$ and consequently we get the following :

Corollary 5.1: *The efficiency fator E of any connected incomplete block design cannot exceed E_0 given by*

$$E_0 = \frac{v(n-b)}{n(v-1)} \quad \dots \quad (5.6)$$

Substituting $n = bk$ in (5.6) we get the following:

Corollary 5.2: *The efficiency factor E of any connected proper incomplete block design cannot exceed E_0 given by*

$$E_0 = \frac{1-1/k}{1-1/v} \quad \dots \quad (5.7)$$

which is the efficiency factor of a BIB design, if one exists for these values of v and k.

A connected incomplete block design with efficiency factor E_0 given in (5.6) will be called a 'most efficient' design.

If the equality in (5.4) is to hold, we must have $n_{jt} = 1$ or 0. If the equality in (5.1) is to hold, the matrix C must have the $(v-1)$ positive latent roots all equal. Also the rows (or columns) of C each adds to zero. It has been shown by Roy and Laha (1956) that the necessary and sufficient condition for a symmetric matrix C of order v to have all diagonal elements equal and all off-diagonal elements equal are that the vector (1, 1, ..., 1) should be a latent vector corresponding to one of the latent roots and that the other $(v-1)$ latent roots should be equal. We thus get the following :

Theorem 5.2 : *The necessary and sufficient condition for an incomplete block design to be most efficient is that the design is binary and its coefficient matrix has all diagonal elements equal and all off-diagonal elements equal.*

It is interesting to note that a BIB design is most efficient, but that Theorem 5.2 does not imply that a most efficient design is necessarily a BIB. However, if we restrict ourselves to proper incomplete block designs, the coefficient-matrix reduces to the form

$$C = R - \frac{1}{k}\Lambda$$

of which the diagonal elements are $\quad c_{jj} = r_j(1-1/k)$

and the off-diagonal elements are $\quad c_{jj'} = -\frac{\lambda_{jj'}}{k}$.

From Theorem 5.2: we now derive the following corollary

Corollary 5.2.1 : *In the class of proper incomplete block designs a most efficient design (if it exists) is necessarily a BIB.*

Another immediate corollary from Theorem 5.2 is the following.

Corollary 5.2.2: *The necessary and sufficient condition for a binary design to be most efficient is that the least squares estimate of the difference between the effects of any two treatments has the same variance.*

6. The efficiency factor of the dual of a given design

In this section we shall be concerned with proper equireplicate binary designs only. Consider one such design D involving v treatments in b blocks each of k cu's with each treatment replicated r times. Let N denote its incidence matrix and $\Lambda = NN'$ its association matrix. The efficiency factor E for this design is given by formula (4.12). The dual of this design, D^* involves $v^* = b$ treatments in $b^* = v$ blocks, each of $k^* = r$ cu's each treatment is replicated $r^* = k$ times. The association matrix of D^* is given by $\Lambda^* = NN'$ and the efficiency factor E by

$$E^* = \frac{v^*-1}{r^*k^*} \bigg/ \left\{ \frac{u^*}{r^*k^*} + \sum_{i=1}^{v^*-u^*-1} \frac{1}{r^*k^*-\phi_i^*} \right\} \qquad \ldots \quad (6.1)$$

where u^* is the number of zero latent roots of Λ^* and ϕ_*'s are the latent roots other than 0 and rk of Λ^*.

Now, the non-zero latent roots of Λ and Λ^* are identical and therefore equating the order of the matrix Λ^* to the total number of its latent roots, we get

$$u^* = b-v+u.$$

On substituting these values in (5.1) we get

$$E^* = \frac{b-1}{rk} \bigg/ \left\{ \frac{b-v+u}{rk} + \sum_{i=1}^{v-u-1} \frac{1}{rk-\phi_i} \right\}. \qquad \ldots \quad (6.2)$$

Using (6.2) and (4.12) we derive the following.

Theorem 6.1: *If E is the efficiency factor of a proper equireplicate binary design involving v treatments in b blocks each of k cu's then the efficiency factor E^* of the dual design is given by*

$$E^* = \frac{(b-1)E}{(b-v)E+(v-1)} \qquad \ldots \quad (6.3)$$

Since E never exceeds unity, we conclude that

$$E^* \gtreqless E \text{ according as } b \gtreqless v. \qquad \ldots \quad (6.4)$$

New designs have sometimes been constructed by dualising well-known designs. The Linked Block (LB) designs obtained by Youden (1951) by dualising BIB designs have been thoroughly investigated by Roy and Laha (1956a) who have found that most of the LB designs with $r, k \leqslant 10$ have efficiency factor of the order of 90 percent. The high efficiency of the LB designs is due to the fact that in a BIB design $b \geqslant v$, so that from (6.4) it follows that the efficiency factor of a LB design must be at least as great as that of the BIB design from which it is dualised. Ramakrishnan (1956) obtained a new class of designs by dualising a number of two-associate PBIB designs with $k \neq 2$. The efficiency factor of these designs can be obtained from formula (6.3).

In general, the overall analysis of variance of the dual of any well-known design can be easily carried out by the 'P-method' suggested by Rao (1956). However, it becomes somewhat cumbersome to compute the variance of different comparisons of the type $\theta_j - \theta_{j'}$. The use of an average variance $\bar{V}$ defined by (4.6) has sometimes been recommended. It is easy to see that $\bar{V}$ for the dual of a design with efficiency factor E and involving v treatments in b blocks each of k eu's with each treatment replicated r times is given by

$$\bar{V} = \frac{2\sigma^2\{(b-v)E+(v-1)\}}{k(b-1)E} \quad \ldots \quad (6.5)$$

σ^2 being the intra-block error. This can therefore be very easily obtained.

REFERENCES

KEMPTHORNE, O. (1956): The efficiency factor of an incomplete block design. *Ann. Math. Stat.*, 27, 847.

RAMAKRISHNAN, C. S. (1956): The dual of a PBIB design and a new class of designs with two replications. *Sankhyā*, 17, 133.

RAO, C. R. (1952): *Advanced Statistical methods in Biometric Research*, John Wiley & Sons, New York.

——— (1955): Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā*, 15, 253.

——— (1956): On the recovery of interblock information in varietal trials. *Sankhyā*, 17, 105.

ROY, J. AND LAHA, R. G. (1956): Classification and analysis of Linked Block designs. *Sankhyā*, 17, 115.

——— (1957): Partially balanced linked block designs, *Ann. Math. Stat.*, 28, 488.

YOUDEN, W. G. (1951): Linked Blocks: A new class of incomplete block designs (Abstract) *Biometrics*, 7, 124.

Paper received : March, 1957.