Jour. Ind. Soc. Ag. Statistics
49 (Golden Jub. No.), 1996-97 : 237–248

Efficiency of the Residual Design Under the Loss of Observations in a Block

Aloke Dey, Chand K. Midha[1] and D.C. Buchthal[1]
Indian Statistical Institute, New Delhi, India

SUMMARY

A situation is considered when an arbitrary number of observations in a single block are lost accidentally from a binary block design. For this, a lower bound to the efficiency factor of the residual design is obtained.

Key words : Missing data, Efficiency factor.

1. Introduction

The robustness of block designs against missing data has been investigated by several authors. For an excellent review of the subject up to 1988, a reference may be made to Kageyama [8] and for more recent references, see Dey [4]. A criterion of robustness of designs was introduced by Ghosh [6], according to which a connected incomplete block design is robust against the loss of $t\,(\geq 1)$ observations if the *residual* design obtained by deleting these t observations remains connected. Following Dey [4], in Section 2, conditions are obtained for a block design to be robust against the loss of an arbitrary number of observations in a block.

The efficiency factor of the residual design when all, or an arbitrary number of observations are lost from a block has been studied for some *specific* incomplete block designs by Srivastava, Gupta and Dey [11], Mukerjee and Kageyama [10], Gupta and Srivastava [7], Das and Kageyama [2], Dey [4] and Duan and Kageyama [5]. In Section 3, a lower bound to the efficiency factor of the residual design is obtained when an arbitrary number of observations are lost from a block of *any* binary block design with constant block size, given that the original design is robust according to the criterion of Ghosh [6]. Some applications are made in Section 4.

Throughout the paper we deal only with real matrices and vectors. Denote an n-component vector of all unities by $\mathbf{1}_n$, an identity matrix of order n by

1 The University of Akron, Akron, Ohio, USA

I_n and an $m \times n$ matrix of all ones by $J_{m,n}$. $J_{m,m}$ is simply denoted by J_m. Further, A', $\mathcal{M}(A)$, A^- and A^+ will respectively denote the transpose, column space (range), a generalized inverse (g-inverse) and the Moore-Penrose inverse of a matrix A. For the definitions of various designs used in the sequel a reference may be made to Dey [3].

2. *Conditions for Robustness*

Let d be a connected, binary block design with v treatments and b blocks each of size k. Suppose t $(1 \le t \le k)$ observations belonging to one of the blocks of d are lost accidentally. Without loss of generality, let the missing observations pertain to the first t treatments in the first block of d. Let C_0 (respectively, C_t) denote the coefficient matrix of the reduced intra-block normal equations under the design d (respectively, the residual design, d_t). Then it is not hard to see that

$$C_0 = C_t + V V' \tag{2.1}$$

where V is a $v \times t$ matrix given by

$$V = \begin{bmatrix} I_t + \beta J_t \\ -F \\ O \end{bmatrix} \tag{2.2}$$

where $\beta = t^{-1}\left(\sqrt{\frac{k-t}{k}} - 1\right)$, $F = \{ k(k-t) \}^{-\frac{1}{2}} J_{k-t,t}$ and O is a $(v-k) \times t$ null matrix. From Theorem 1 of Dey [4], it follows that the design d is robust against the loss of t observations $(1 \le t \le k)$ in a block if and only if $I_t - V' C_0^- V$ is positive definite.

Remark 2.1. We note in passing that the condition $\mathcal{M}(G) \subset \mathcal{M}(A)$ in Theorem 1 of Dey [4] is redundant. In fact, if A and B are a pair of real symmetric nonnegative definite matrices such that $A - B$ is also nonnegative definite, then $\mathcal{M}(B) \subset \mathcal{M}(A)$. This can be seen as follows : write $A = B + C$, where C is symmetric nonnegative definite. Then there exist matrices X and Y such that $B = X X'$ and $C = Y Y'$. Hence

$$A = X X' + Y Y' = [\, X \; Y \,] \begin{bmatrix} X' \\ Y' \end{bmatrix}$$

which shows that $\mathcal{M}(B) = \mathcal{M}(X X') = \mathcal{M}(X) \subset \mathcal{M}(A)$.

The following result is not difficult to prove.

Lemma 2.1. With V as in (2.2),

(i) $1'_v V = 0'$

(ii)
$$VV' = \begin{bmatrix} I_t - k^{-1} J_t & -k^{-1} J_{t,k-t} & O \\ -k^{-1} J_{k-t,t} & tJ_{k-t} / \{k(k-t)\} & O \\ O & O & O \end{bmatrix}, \text{ if } 1 \le t \le k-1$$

$$= \begin{bmatrix} I_k - k^{-1} J_k & O \\ O & O \end{bmatrix}, \text{ if } t = k$$

(iii) $V V'$ is idempotent for $1 \le t \le k$ and

$$\text{Rank}(V V') = t, \text{ if } 1 \le t \le k-1$$
$$= k-1, \text{ if } t = k$$

(iv) $V' V = I_t$, if $1 \le t \le k-1$

$$= I_k - k^{-1} J_k, \text{ if } t = k$$

Since $V V'$ is idempotent, following Dey [4] we obtain the following sufficient condition for the robustness of d.

Theorem 2.1. The design d is robust against the loss of $t\,(1 \le t \le k)$ observations in a block if the smallest positive eigenvalue of C_0 is larger than unity.

As mentioned by Dey [4], the following designs satisfy the condition of Theorem 2.1 :

(i) All balanced incomplete block designs; (ii) all group-divisible designs with the exception of the design with parameters $v = 4 = b$, $r = k = 2$, $m = 2 = n$, $\lambda_1 = 0$, $\lambda_2 = 1$; (iii) all triangular designs with the exception of the design with parameters $v = 10$, $b = 15$, $r = 3$, $k = 2$, $\lambda_1 = 0$, $\lambda_2 = 1$; (iv) all Latin-square type PBIB designs with the exception of L_2 designs with parameters $v = s^2$, $b = 2s$, $r = 2$, $k = s$, $\lambda_1 = 1$, $\lambda_2 = 0$; (v) all PBIB designs based on partial geometries with more than two replicates. Thus the class of designs satisfying the condition of Theorem 2.1 is quite rich.

3. A Lower Bound to the Efficiency Factor

In this section a lower bound to the efficiency factor of the residual design is given when an arbitrary number of observations are lost from a block of a connected, equi-block sized, binary block design. We restrict attention to those designs {d} which satisfy the sufficient condition of Theorem 2.1. As a measure of efficiency of the residual design, we take the quantity E, given by $E = \text{tr}(C_0^+) / \text{tr}(C_t^+)$, where tr (·) stands for the trace of a square matrix. Note that since the trace of C_0^+ (respectively, C_t^+) is equal to the sum of the reciprocals of positive eigenvalues of C_0 (respectively, C_t), this measure of efficiency is related to the well known A—efficiency criterion.

We first prove the following lemma.

Lemma 3.1. Let A be a nonnegative definite matrix of order v and rank $v-1$ and B be a $v \times t$ matrix such that BB' is idempotent of rank $t < v$. Further, let $I_t - B'AB$ and $I_v - A$ be both positive definite. Then,

$$\text{tr}[AB(I_t - B'AB)^{-1}B'A] \leq \sum_{i=1}^{t} \phi_i^2/(1-\phi_i)$$

where $1 > \phi_1 \geq \phi_2 \geq \ldots \geq \phi_t \ldots \geq \phi_{v-1}$ are the positive eigenvalues of A.

Proof. Recall that if $I_n - D$ is a positive definite matrix of order n, then

$$(I_n - D)^{-1} = \sum_{j=0}^{\infty} D^j, \quad \text{where} \quad D^0 = I_n.$$

Now,

$$\text{tr}[AB(I_t - B'AB)^{-1}B'A] = \text{tr}[B(I_t - B'AB)^{-1}B'A^2]$$

$$= \text{tr}[B \sum_{j=0}^{\infty} (B'AB)^j B'A^2]$$

Let H be an orthogonal matrix such that

$$HBB'H' = \begin{bmatrix} I_t & O \\ O & O \end{bmatrix}$$

Define $B_1 = H B H'$ and $A_1 = H A H'$. Then, it is easy to see that

$$\operatorname{tr}[B' \sum_{j=0}^{\infty} (B'AB)^j B' A^2] = \operatorname{tr}[B'_1 \sum_{j=0}^{\infty} (B'_1 A_1 B_1)^j B'_1 A_1^2]$$

It can be seen that for a fixed j,

$$B_1 (B'_1 A_1 B_1)^j B'_1 = \begin{bmatrix} A_{11}^j & O \\ O & O \end{bmatrix} = C \text{ (say)}$$

where

$$A_1 = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and A_{11} is $t \times t$.

Also, it is known (see e.g., Marshall and Olkin ([9], p. 248)) that if U and V are $n \times n$ nonnegative definite (real) symmetric matrices, then $\operatorname{tr}(UV) \leq \sum_{i=1}^{n} \lambda_i(U) \lambda_i(V)$, where for a (real) symmetric nonnegative definite matrix A, $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A)$ denote the eigenvalues of A. Applying this result we have

$$\operatorname{tr}[B'_1 \sum_{j=0}^{\infty} (B'_1 A_1 B_1)^j B'_1 A_1^2] = \sum_{j=0}^{\infty} \operatorname{tr}(C A_1^2)$$

$$\leq \sum_{j=0}^{\infty} \sum_{i=1}^{t} \lambda_i(A_{11}^j) \lambda_i(A_1^2)$$

$$\leq \sum_{j=0}^{\infty} \sum_{i=1}^{t} \lambda_i(A_1^j) \lambda_i(A_1^2), \text{ by the interlacing property}$$

$$= \sum_{j=0}^{\infty} \sum_{i=1}^{t} \lambda_i(A_1^{j+2})$$

$$= \sum_{i=1}^{t} \lambda_i(A_1^2) \sum_{j=0}^{\infty} \lambda_i(A_1^j)$$

$$= \sum_{i=1}^{t} \lambda_i (A_1^2) / (1 - \lambda_i (A_1))$$

The proof is completed by noting that $\lambda_i (A_1) = \phi_i$ for $i = 1, 2, \ldots, v-1$.

From Theorem 2 of Dey [4] we have

$$\operatorname{tr}(C_t^+) = \operatorname{tr}(C_0^+) + \operatorname{tr}[C_0^+ V (I_t - V' C_0^+ V)^{-1} V' C_0^+]$$

where V is as in (2.2). Using Lemma 3.1 we have

$$E^{-1} = 1 + \frac{\operatorname{tr}[C_0^+ V (I_t - V' C_0^+ V)^{-1} V' C_0^+]}{\operatorname{tr}(C_0^+)}$$

$$\leq 1 + \frac{\sum_{i=1}^{t} \xi_i^2 / (1 - \xi_i)}{\sum_{i=1}^{v-1} \xi_i}$$

where $1 > \xi_1 \geq \xi_2 \geq \ldots \geq \xi_{v-1}$ are the positive eigenvalues of C_0^+. Recall that since we are considering only those designs that satisfy the condition of Theorem 2.1, we have for $i = 1, 2, \ldots, v-1$; $\xi_i < 1$.

If $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{v-1}$ are the positive eigenvalues of C_0, then $\theta_i > 1$, for $i = 1, 2, \ldots, v-1$, since the positive eigenvalues of C_0^+ are the reciprocals of the positive eigenvalues of C_0. We therefore have the following result.

Theorem 3.1. A lower bound to the efficiency factor of the residual design d_t is given by

$$E \geq \left[1 + \frac{\sum_{i=1}^{t} \theta_i^{-1} (\theta_i - 1)^{-1}}{\sum_{i=1}^{v-1} \theta_i^{-1}} \right]^{-1}$$

where $1 < \theta_1 \leq \theta_2 \leq \ldots \leq \theta_{v-1}$ are the positive eigenvalues of C_0.

4. Applications

In this section specific results on the efficiency factor of the residual design are given when an arbitrary number of observations are lost from a block in the case of some important classes of designs. Throughout this section, we consider only those designs that satisfy the sufficient condition of Theorem 2.1.

4.1 *Balanced incomplete block designs*

Let d be a balanced incomplete block (BIB) design with usual parameters v, b, r, k, λ. For such a design one can evaluate the *exact* efficiency factor of the residual design, without computing the eigenvalues of C_t.

For a BIB design the matrix C_0 is given by

$$C_0 = \frac{\lambda v}{k}(I_v - v^{-1} J_v)$$

and hence $C_0^+ = (k/(\lambda v))(I_v - v^{-1} J_v)$. Now we distinguish two cases according as $1 \le t \le k-1$ or $t = k$. When $1 \le t \le k-1$ we have, by letting $\alpha = k/(\lambda v)$,

$$\begin{aligned} \mathrm{tr}(C_t^+) &= \mathrm{tr}(C_0^+) + \mathrm{tr}[C_0^+ V (I_t - V' C_0^+ V)^{-1} V' C_0^+] \\ &= (v-1)\alpha \\ &\qquad + \mathrm{tr}[\alpha^2 (I_v - v^{-1} J_v) V (I_t - \alpha V' (I_v - v^{-1} J_v) V)^{-1} V' (I_v - v^{-1} J_v)] \\ &= (v-1)\alpha + \mathrm{tr}[\alpha^2 V (I_t - \alpha I_t)^{-1} V'], \text{ by Lemma 2.1} \\ &= (v-1)\alpha + \frac{\alpha^2}{1-\alpha}\mathrm{tr}(VV') \\ &= (v-1)\alpha + \frac{\alpha^2 t}{(1-\alpha)} \end{aligned}$$

Plugging in the value of α and simplifying we obtain

$$E = (v-1)(\lambda v - k) / \{(v-1)(\lambda v - k) + tk\}, \quad 1 \le t \le k-1$$

This expression was obtained earlier by Das and Kageyama [2] by actually computing the eigenvalues of C_t.

For $t = k$ we have, since $V'V = I_k - k^{-1} J_k$

$$\begin{aligned}
\mathrm{tr}\,(C_k^+) &= \mathrm{tr}\,(C_0^+) + \mathrm{tr}\,[\,C_0^+ V (I_k - V' C_0^+ V)^{-1} V' C_0^+\,] \\
&= (v-1)\,\alpha + \alpha^2\,\mathrm{tr}\,[V (I_k - \alpha V'V)^{-1} V'\,] \\
&= (v-1)\,\alpha + \alpha^2 (1-\alpha)^{-1}\,\mathrm{tr}\,[VV'], \\
&\qquad \text{as } (I_k - \alpha V'V)^{-1} = (1-\alpha)^{-1} (I_k - (\alpha/k) J_k) \\
&= (v-1)\,\alpha + \alpha^2 (1-\alpha)^{-1} (k-1)
\end{aligned}$$

Putting the value of α and simplifying gives

$$E = (v-1)\;(\lambda v - k) / \{\,(v-1)(\lambda v - k) + k(k-1)\},\quad t = k$$

Once again this expression agrees with the one given by Das and Kageyama [2]. Note that if for $1 \le t \le k$, E_t denotes the efficiency factor when t observations in a block are lost then E_t is a monotonically decreasing function of t and $E_k = E_{k-1}$.

4.2 *Group-divisible designs*

Consider now a group-divisible design with usual parameters $v = mn, b, r, k, \lambda_1, \lambda_2, m, n$. The matrix C_0 in this case has two positive eigenvalues, $\theta_1 = \{\,r(k-1) + \lambda_1\,\}/k$ with multiplicity $v - m$ and $\theta_2 = v\lambda_2/k$ with multiplicity $m - 1$. It is easy to see that $\theta_1 < \theta_2$ if and only if $\lambda_1 < \lambda_2$. Observe that no singular group-divisible design can have $\lambda_1 < \lambda_2$, while all semi-regular group-divisible designs satisfy $\lambda_1 < \lambda_2$. Further, it is easy to see that for a semi-regular group-divisible design, $v - m \ge k$. For regular group-divisible designs satisfying $\lambda_1 < \lambda_2$ it is not necessary that $v - m \ge k$, though a large number of designs listed by Clatworthy [1] do have this property. In fact, among the regular group-divisible designs listed by Clatworthy, there are 70 designs that satisfy $\lambda_1 < \lambda_2$ and among these 70 designs, there are only 5 designs for which $v - m < k$. These designs are R96, R134, R136, R175 and R205. If we restrict attention to all group-divisible designs satisfying $\lambda_1 < \lambda_2$ and $v - m \ge k$, then one can take $t \le k \le v - m$. Letting

$$\gamma = \frac{\theta_1}{(\theta_2 - 1)\,\{\,(v-m)\,\theta_2 + (m-1)\,\theta_1\,\}}$$

$$\delta = \frac{\gamma\theta_2 (\theta_2 - 1)}{\theta_1 (\theta_1 - 1)}$$

and using Theorem 3.1 we arrive at the following result.

Theorem 4.1. For a group-divisible design with parameters $v = mn$, b, r, k, λ_1, λ_2, m, n satisfying $\lambda_1 < \lambda_2$ and $v - m \geq k$, the efficiency factor E of the residual design has a lower bound E_0 given by

$$E_0 = (1 + t\delta)^{-1}, \ 1 \leq t \leq k \tag{4.1}$$

If in a group-divisible design $\theta_1 > \theta_2$ (equivalently, $\lambda_1 > \lambda_2$), then we have either $k \leq m - 1$ or $k > m - 1$. For each of these cases, results on the bound E_0 are summarized below.

Theorem 4.2. For a group-divisible design with parameters $v = mn, b, r, k, \lambda_1, \lambda_2, m, n$, satisfying $\lambda_1 > \lambda_2$, the efficiency factor E of the residual design has a lower bound E_0 given by

$$E_0 = (1 + t\gamma)^{-1}, \text{ if } k \leq m - 1; \ 1 \leq t \leq k \tag{4.2}$$

$$= (1 + t\gamma)^{-1}, \text{ if } k > m - 1, \ 1 \leq t \leq m - 1 \tag{4.3}$$

$$= \{ 1 + (m - 1)\gamma + (t - m + 1)\delta \}^{-1}, \text{ if } t \geq m \tag{4.4}$$

Clearly, in each of the cases under Theorems 4.1 and 4.2, E_0 is a monotonically decreasing function of t, as expected. The values of E_0 given by Theorems 4.1 and 4.2 were computed for all group-divisible designs listed in Clatworthy [1] for $1 \leq t \leq k$. Let the minimum of E_0 be E^*, i.e., E^* is the value of E_0 for $t = k$. The results of the computations show that 5 designs have $E^* \leq 0.50$, 16 designs have $0.50 < E^* \leq 0.70$, 31 designs have $0.70 < E^* \leq 0.80$, 346 designs have $0.80 < E^* \leq 0.95$ and 42 designs have $E^* > 0.95$. This shows that only in about 12% cases, the loss in efficiency could be 20% or more when all observations are lost in a block of a group divisible design.

Mukerjee and Kageyama [10] obtained lower and upper bounds on the efficiency factor of the residual design when *all* observations in a block of a regular group-divisible design satisfying $\lambda_1 > 0$ are lost. Their upper bound is simply $(b - 1)/b$. A comparison of E^* with the lower bound of Mukerjee and Kageyama shows that for 27 designs, our bound is either equal to or greater than the Mukerjee-Kageyama lower bound. The designs for which our bound is sharper than the Mukerjee-Kageyama lower bound are: R32, R53, R74, R102, R127, R135, R166, R168, R173, R178, R182, R187, R188, R195, R198, R204, R206, R207 and R208.

4.3 *Triangular designs*

Consider now a triangular design with parameters $v = n(n-1)/2, b, r, k, \lambda_1, \lambda_2$, where $n \geq 5$ is an integer. The positive eigenvalues of C_0 in this case are

$$\theta_1 = k^{-1}\{n\lambda_1 + n(n-3)\lambda_2/2\}$$

and

$$\theta_2 = k^{-1}\{2(n-1) + \lambda_2(n-1)(n-4)/2\}$$

with respective multiplicities $n-1$ and $n(n-3)/2$. Thus, $\theta_1 < \theta_2$ if and only if $\lambda_1 > \lambda_2$. From Theorem 3.1 we therefore arrive at the following result.

Theorem 4.3. (a) For a triangular design with parameters $v = n(n-1)/2, b, r, k, \lambda_1, \lambda_2$ satisfying $\lambda_1 < \lambda_2$, the efficiency factor of the residual design has a lower bound E_0 given by

$$E_0^{-1} = 1 + 2t\theta_2/\delta, \text{ if } t \leq n-1$$
$$= 1 + \{2(n-1)(\theta_2 - \theta_1) + 2t\theta_1\}/\delta \quad \text{if } t > n-1$$

(b) For a triangular design with parameters $v = n(n-1)/2$, b, r, k, λ_1, λ_2 satisfying $\lambda_1 > \lambda_2$, the efficiency factor of the residual design has a lower bound E_0 given by

$$E_0^{-1} = 1 + 2t\theta_1/\delta, \text{ if } t \leq n(n-3)/2$$
$$= 1 + \{n(n-3)(\theta_1 - \theta_2) + 2t\theta_2\}/\delta \text{ if } t > n(n-3)/2$$

where $\delta = 2(n-1)\theta_2 + n(n-3)\theta_1$

The values of E_0 given by Theorem 4.3 were computed for all the triangular designs in the catalogue of Clatworthy [1], for $1 \leq t \leq k$. The design with parameters $v = 10$, $b = 15$, $r = 3$, $k = 2$, $\lambda_1 = 0$, $\lambda_2 = 1$ was left out, as for this design the condition of Theorem 2.1 is not met. (By direct computation it is seen that for this design $E^* = 0.85$). The results of this computation reveals that for 34 designs, $0.50 < E^* \leq 0.70$, 42 designs have $0.70 < E^* \leq 0.80$, 20 designs have $0.80 < E^* \leq 0.95$ and two designs have $E^* > 0.95$, where E^* as before denotes the minimum value of E_0. This shows that for a large number of triangular designs, the loss in efficiency when all observations in a block are lost could be 20% or more.

Remark 4.1. The lower bound to the efficiency factor given by Theorem 3.1 may not always be sharp. This however is expected as no design structure has been assumed in deriving the bound. On the other hand, the bound is applicable to *any* binary equi-block sized design.

ACKNOWLEDGEMENTS

The authors wish to thank a referee for very useful comments on previous versions and Professor R.B. Bapat for helpful discussions.

REFERENCES

[1] Clatworthy, W.H., 1973. *Tables of Two-Associate-Class Partially Balanced Designs*. Nat. Bur. Standards, Appl. Math. Ser. No. 63, Washington, D.C.

[2] Das, A. and Kageyama, S., 1992. Robustness of BIB and extended BIB designs against the nonavailability of any number of observations in a block. *Comput. Statist. Data Anal.*, **14,** 343-358.

[3] Dey, A., 1985. *Theory of Block Designs*. Halsted, New York.

[4] Dey, A., 1993. Robustness of block designs against missing data. *Statistica Sinica*, **3,** 219-231.

[5] Duan, Xiaoping and Kageyama, S., 1995. Robustness of augmented BIB designs against the unavailability of some observations. *Sankhya*, **B57,** 405-419.

[6] Ghosh, S., 1982. Robustness of BIBD against the unavailability of data. *J. Statist. Plann. Inf.*, **6,** 29-32.

[7] Gupta, V.K. and Srivastava, R., 1992. Investigation of robustness of block designs against missing observations. *Sankhya*, **B54,** 100-105.

[8] Kageyama, S., 1990. Robustness of block designs. *Probability, Statistics and Design of Experiments* (Edited by R.R. Bahadur), 425 - 438, Wiley Eastern, New Delhi.

[9] Marshall, A.W. and Olkin, I., 1979. *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York.

[10] Mukerjee, R. and Kageyama, S., 1990. Robustness of group-divisible designs. *Comm. Statist.—Theory Methods,* **19**, 3189-3203.

[11] Srivastava, R., Gupta, V.K. and Dey, A., 1990. Robustness of some designs against missing observations. *Comm. Statist.—Theory Methods,* **19**, 121-126.