

## ON ROBUSTNESS OF DESIGNS AGAINST INCOMPLETE DATA

By SUBIR GHOSH

Indian Statistical Institute

**SUMMARY.** In this paper, we characterize the robustness property of designs against incomplete data in the sense that, when any  $t$  (a positive integer) observations are missing, all parameters are still estimable in the model assumed. We also present some examples of Srivastava-Chopra Optimum balanced resolution  $V$  plans for  $2^m$  factorials which are robust against missing of any two observations.

### 1. INTRODUCTION

The robustness of designs against incomplete data in case of missing of any single observation was first considered in Ghosh (1978). This paper gives a characterization of robustness property in the general case of missing of any  $t$  observations. Some examples of designs robust against missing of any two observations are also presented.

### 2. ROBUST DESIGNS

Consider the ordinary linear model

$$E(y) = A\xi \quad \dots (1)$$

$$V(y) = \sigma^2 I_N \quad \dots (2)$$

$$\text{Rank } A = \nu \quad \dots (3)$$

where  $y(N \times 1)$  is a vector of observations,  $A(N \times \nu)$  is a known matrix,  $\xi(\nu \times 1)$  is a vector of fixed unknown parameters and  $\sigma^2$  is a constant which may or may not be known. Let  $T$  be the underlying design corresponding to  $y$ .

**Definition 1:** A design under the model (1-3) is said to be robust against missing of any  $t$  (a positive integer) observations if the  $(N-t \times \nu)$  matrix obtained from  $A$  by omitting any  $t$  rows has rank  $\nu$ . It is clear from definition 1 that  $N$  must at least be  $\nu+t$ . Suppose  $N = \nu+k$ , where  $k (> t)$  a positive integer. Clearly, there exist  $k$  linearly independent vectors  $C_i' = (C_{i1}, \dots, C_{iN})$ ,  $i = 1, \dots, k$ , with real elements satisfying

$$C_i' A = 0 \quad \dots (4)$$

Consider the  $(k \times N)$  matrix

$$C = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1t} & \dots & C_{1N} \\ C_{21} & C_{22} & \dots & C_{2t} & \dots & C_{2N} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ C_{k1} & C_{k2} & \dots & C_{kt} & \dots & C_{kN} \end{bmatrix} \quad \dots \quad (5)$$

whose  $i$ -th row is  $C_i$  and furthermore,  $\text{Rank } C = k$ . We now recall that a matrix  $B$  is said to have the property  $P_t$  if no  $t$  columns of  $B$  are linearly dependent. The following theorem characterizes the robustness property.

**Theorem 1:** *Let  $T$  be a design under (1-3) with  $N = v + k$  observations, where  $k (\geq t)$  a positive integer. Then,  $T$  is robust against missing of any  $t$  observations if and only if (iff) the matrix  $C$ , defined in (5), has the property  $P_t$ .*

*Proof:* Suppose  $C$  has  $P_t$ . Let

$$A = \begin{bmatrix} A_1 \\ \dots \\ A_2 \end{bmatrix}, \quad C = [C_1^* : C_2^*], \quad \dots \quad (6)$$

where  $A_1(t \times v)$ ,  $A_2(\overline{N-t} \times v)$ ,  $C_1^*(k \times t)$  and  $C_2^*(k \times \overline{N-t})$ .

We have, from (4),

$$C_1^* A_1 + C_2^* A_2 = 0. \quad \dots \quad (7)$$

Suppose

$$C_1^* = \begin{bmatrix} C_{11}^* \\ \dots \\ C_{1t}^* \end{bmatrix}, \quad C_2^* = \begin{bmatrix} C_{21}^* \\ \dots \\ C_{2t}^* \end{bmatrix} \quad \dots \quad (8)$$

where  $C_{11}^*(t \times t)$ ,  $C_{1t}^*(\overline{k-t} \times t)$ ,  $C_{21}^*(t \times \overline{N-t})$ , and  $C_{2t}^*(\overline{k-t} \times \overline{N-t})$ . Suppose, furthermore,  $\text{Rank}(C_{11}^*) = t$ . Thus, we get

$$C_{11}^* A_1 + C_{21}^* A_2 = 0. \quad \dots \quad (9)$$

Hence,

$$A_1 = -C_{11}^{*^{-1}} C_{21}^* A_2. \quad \dots \quad (10)$$

Thus, the rows of  $A_1$  are linear combinations of the rows in  $A_2$ . Therefore, the matrix  $A_2$  obtained from  $A$  by omitting  $t$  rows in  $A_1$ , has rank  $v$ . The argument is similar for any other set of  $t$  rows of  $A$ . Hence the design  $T$  is robust.

Suppose the design  $T$  is robust against missing of  $t$  observations. Then, there is a  $(t \times \sqrt{N-t})$  matrix  $D$  satisfying

$$A_1 = DA_2 \quad \dots \quad (11)$$

$$\text{i. e.,} \quad [I_t : -D] \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = 0. \quad \dots \quad (12)$$

Considering (4), (6), and (9), it follows that there exists a  $(t \times k)$  matrix  $U$  such that

$$UC_1^* = I_t, \quad UC_2^* = -D. \quad \dots \quad (13)$$

It is now easy to check that  $\text{Rank}(C_1^*) = t$ . Therefore  $C$  has  $P_t$ . This completes the proof of the theorem.

The following results are of practical importance.

Corollary 1: Suppose  $t = 1$ . The design  $T$  is robust against missing of any one observation iff  $C(k \times N)$  has the property  $P_1$  or, in other words,

$$(C_{1j}, C_{2j}, \dots, C_{kj}) \neq (0, 0, \dots, 0) \text{ for } (j = 1, \dots, N)$$

(i.e., none of the column vectors in  $C$  is a null vector).

Corollary 2: Suppose  $t = 2$ . The design  $T$  is robust against missing of any two observations iff  $C(k \times N)$  has the property  $P_2$ , or in other words,

$$(i) \quad (C_{1j}, C_{2j}, \dots, C_{kj}) \neq (0, 0, \dots, 0) \text{ for } (j = 1, \dots, N),$$

$$(ii) \quad (C_{1j}, C_{2j}, \dots, C_{kj}) \neq w(C_{1j'}, C_{2j'}, \dots, C_{kj'}),$$

where  $j \neq j'$ ,  $(j, j' = 1, \dots, N)$ , and  $w$  is a real constant.

It is to be remarked that the above results are also true in case  $A(N \times M)$ ,  $\xi(M \times 1)$  and  $\text{Rank}(A) = v < \min(M, N)$ .

### 3. EXAMPLES FROM $2^m$ FACTORIALS

Consider a  $2^m$  factorial experiment. The treatments are denoted by  $(x_1, x_2, \dots, x_m)$ , where  $x_t = 0$  or 1. We denote a design with  $N$  treatments by a  $(N \times m)$  matrix  $T$  whose rows are treatments. Optimal balanced resolution  $V$  plans for  $2^m$  factorials,  $4 \leq m \leq 8$ , and for practical values of  $N$ , have been presented in the papers of Srivastava and/or Chopra.

By 'weight' of a vector, we mean the number of nonzero elements in it. Let  $S_i$  be the set of all  $(1 \times m)$  vectors, with elements 0 and 1, of weight  $i$  ( $i = 0, 1, \dots, m$ ). Clearly the number of members in  $S_i$  is  $\binom{m}{i}$ .

Srivastava-Chopra designs are denoted by  $\lambda' = (\lambda_0, \lambda_1, \dots, \lambda_m)$  where  $\lambda_i$  is the number of times the set  $S_i$  occurs in the design. Thus  $N = \sum_{i=0}^m \binom{m}{i} \lambda_i$ . These optimum designs may or may not remain optimum or even resolution  $V$  plans when some observations are missing. We now present, as example, designs which are robust against missing of any  $t$  observations. These designs remain as resolution  $V$  plans when any  $t$  observations are missing.

*Example 1:* Consider  $m = 4, N = 15$ . Here,  $\nu = 11$ . Thus  $k = 4$ . The design is represented as  $\lambda' = (1 \ 1 \ 1 \ 1 \ 0)$ . The matrix  $C$  is given below

$$C = \begin{bmatrix} 1 & -3 & -3 & -3 & -3 & 2 & 2 & 2 & 2 & 2 & 2 & -1 & -1 & -1 & -1 \\ 0 & 1 & -1 & 1 & -1 & 0 & -2 & 0 & 0 & 2 & 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & -1 & 1 & 0 & 0 & -2 & 2 & 0 & 0 & -1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 & -2 & 0 & 0 & 0 & 0 & 2 & 1 & 1 & -1 & -1 \end{bmatrix}$$

It is easy to check that the above matrix has the property  $P_2$  but not  $P_3$ . Thus the present design is robust against missing of any two observations and not robust against missing of any three observations. Clearly for  $N = 16$ , the design  $\lambda' = (1 \ 1 \ 1 \ 1 \ 1)$  is also robust against missing of two observations.

*Example 2:* Consider  $m = 5, N = 22a$ . We have  $\nu = 16$  and thus ( $k = 6$ ). The design is given by  $\lambda' = (1 \ 1 \ 1 \ 0 \ 1 \ 1)$ . We present the matrix  $C$  below

$$C = \begin{bmatrix} 3 & -3 & -3 & -3 & -3 & -3 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & -3 & -3 & -3 & -3 & 7 \\ -3 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 3 \\ 0 & -2 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & -1 & -1 & -1 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 2 & 0 & 0 & 1 & 0 & 1 & 1 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & -2 & 0 & 0 & 2 & 0 & 1 & 1 & 0 & 1 & 0 & -1 & 0 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 2 & 1 & 1 & 1 & 0 & 0 & -1 & 0 & -1 & -1 & 1 & 0 & 0 & -1 & 0 \end{bmatrix}$$

Observe that the above matrix has the property  $P_2$  and, therefore, this design is robust against missing of any two observations. It is clear that the designs  $N = 23a, \lambda' = (2 \ 1 \ 1 \ 0 \ 1 \ 1)$ ,  $N = 24a, \lambda' = (2 \ 1 \ 1 \ 0 \ 1 \ 2)$ , and  $N = 25a, \lambda' = (3 \ 1 \ 1 \ 0 \ 1 \ 2)$  have the same property.

## REFERENCES

- BOSE, R. C. (1947): Mathematical theory of symmetrical factorial designs. *Sankhy* 8, 107-88.
- BOX, G. E. P. and DRAPER, N. R. (1975): Robust designs. *Biometrika*, 62, 2, 347-52.
- GHOSH, S. (1978): On robustness of optimal balanced resolution  $V$  plans. (Submitted for publication)
- HERZBERG, A. M. and ANDREWS, D. F. (1976): Some considerations in the optimal design of experiments in non-optimal situations. *J. Roy. Stat. Soc. Ser. B*, 3, 284-89.
- HUBER, P. (1975): *Robustness and Designs. A Survey of Statistical Design and Linear Models* (Ed: J. N. Srivastava), Amsterdam: North-Holland Publishing Company, 287-301.
- SRIVASTAVA, J. N. and CHOPRA, D. V. (1971): Balanced optimal  $2^m$  fractional factorial designs of resolution  $V$ ,  $m < 6$ . *Technometrics*, 13, 257-69.

*Paper received: October, 1978.*

*Revised: January, 1979.*