Jour. Ind. Soc. Ag. Statistics
56(1), 2003 : 7-18

Robust Block Designs for Making Test Treatments-Control Treatment Comparisons Against the Presence of an Outlier

S. Sarker, V.K. Gupta and Rajender Parsad
Indian Agricultural Statistics Research Institute, New Delhi–110 012
(Received : February, 2002)

SUMMARY

In the present investigation the general expression of Cook-statistic has been derived for detection of outlier in the block designs for making test treatments-control treatment comparisons. The criterion of minimization of average Cook-statistic has been modified to identify robust designs for making test treatments-control treatment comparisons. It has been shown that all BTIB designs that are binary with respect to test treatments are robust against the presence of a single outlier.

Key words: BTIB design, Outlier, Cook-statistic, Robustness.

1. Introduction

The problem of outliers in the data generated from a general linear model has been studied extensively in the literature. Among various statistics developed for the detection of outliers, Cook's distance developed by Cook [4] has been extensively used. Outliers are likely to occur in the data generated from experimental designs due to disease and / or insect attack on some particular plot of the experiment, heavy irrigation by mistake on some particular block(s) or plot(s) of the experiment, mistakes creeping in during recording of data, etc. In the data generated from designed experiments, the problem of outlier(s), however, has not been studied so extensively, though there are some references available, see e.g., Box and Draper [3], Gopalan and Dey [9], Singh *et al.* [13], Ghosh ([6], [7]) and Ghosh and Kipngeno [8]. Box and Draper [3] were the first to study the robustness of experimental designs in presence of outlier(s). Box and Draper suggested that in order to make a designed experiment insensitive to outlier, the variance of overall discrepancy in the predicted responses should be minimised. Ghosh and Kipngeno [8] used the same criterion for studying the robustness of optimum balanced 2^m factorial designs of resolution. V. Ghosh [6] used the information contained in a set of observations and Cook's distance for detecting the influential observations under full rank models. Ghosh [7] gave another measure of identifying the influential set of observations. The criterion is based on the sum of the variances of the predicted values of the set of unavailable observations and showed that this criterion is equivalent to the Cook's distance. Most of these criteria are based on full rank model situations and can usefully be employed in fitting response surfaces. However, these cannot be applied to the linear model for which the design matrix is deficient in rank. Gopalan and Dey [9] developed a criterion of robustness on the lines similar to those given by Box and Draper [3] in other experimental situations where the design matrix is not of full rank. Instead of taking the overall discrepancy in the variance of estimated value of responses they considered the discrepancy in the estimation of error variance. Gopalan and Dey [9] studied the robustness aspect of experimental designs by minimizing the variance of the bias or discrepancy in the estimation of error variance. They have identified some robust designs for one-way elimination of heterogeneity settings. Singh *et al.* [13] extended the results of Gopalan and Dey [9] to find out robust designs for two-way elimination of heterogeneity setting. They showed that important class of variance balanced row-column designs that satisfy the property of adjusted orthogonality are robust in the presence of a single outlier. Bhar and Gupta [2] have investigated the problem of outliers in the experimental data for the block designs using Cook's distance for the problem of inferring on the complete set of linearly independent orthonormalized treatment contrasts. However, there do exist experimental situations where the interest of the experimenter is only in a subset of all the possible elementary treatment contrasts rather than the complete set of all the possible elementary contrasts. Such an experimental situation could be one where the experimenter is interested in comparing several new treatments called test treatments with one or more standard treatments called control treatments. For example, in genetic resources environment, an essential activity is to test or evaluate the new germplasm / provenance / superior selections (test treatments), etc. with the existing provenance or released varieties (control treatments). National Bureau of Plant Genetic Resources, New Delhi, India, conducts several such trials. For example, a Coordinated advanced trial - I (1992) was conducted to evaluate 15 genotypes of Moth Bean (test treatments), *viz.*, RMO-96, RMO-173, RMO-224, RMO-225, RMO-226, RMO-256, JMS-1, JMS-2, JMS-3, JMS-4, JMM-259, IPCMO-371, IPCMO-481, IPCMO-526, IPCMO-880 with three control treatments, *viz.*, Jadia, Jawala and Maru Moth-I. The trial was conducted in a randomized complete block design to compare each of the test treatments with control treatments. Similar situations also occur in other disciplines of agricultural sciences, industry, etc. The main problem of interest here is to design an experiment for the estimation of test treatments vs control treatments contrasts with as much precision as possible when the comparisons among the test treatments or among the controls are of lesser consequence. It is well known that for the estimation of elementary contrasts among test treatments and control treatments, conventional designs like balanced incomplete block (BIB) designs are not the best. A lot of literature is available for obtaining efficient block designs for making test treatments - control treatments comparisons. For an excellent review on the topic one may refer to Hedayat *et al.* [11], Majumdar [12] and Gupta and Parsad [10] and the references cited therein.

Outliers are likely to occur in the experimental settings just described. However, for the experimental settings where one is interested only in a subset of all the possible elementary treatment contrasts rather than the complete set of all possible elementary treatment contrasts, the problem of outlier(s) needs special attention and has to be handled separately. So the present paper deals with the problem of studying outliers in designed experiments for experimental setting where several test treatments and a single control treatment are run in a design and the interest of the experimenter is to estimate only test treatments vs control treatment contrasts. We begin with some preliminaries in Section 2.

2. *Preliminaries*

Consider an experimental situation where it is required to compare v test treatments with a single control treatment via n experimental units arranged in b blocks, where the j^{th} block contains k_j experimental units $\forall$ j = 1, 2, ..., b, such that $\sum_{j=1}^{b} k_j = n$. The v test treatments are indexed as 1, 2,..., v and the control treatment as 0. It is also assumed that $k_j < (v + 1)$ and that only one treatment is applied to each experimental unit. In design d the i^{th} treatment is applied in n_{dij} experimental units in the j^{th} block. The observations are represented by a two-way classified additive, fixed effects, linear, homoscedastic model

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{\Delta}'\boldsymbol{\tau} + \mathbf{D}'\boldsymbol{\beta} + \mathbf{e} \tag{2.1}$$

that can be rewritten as a general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e} \tag{2.2}$$

with E(**e**) = **0** and V(**e**) $= \sigma^2\mathbf{I}_n$. Here **y** is an n × 1 vector of observations, $\mathbf{X} = [\mathbf{1} \ \ \mathbf{\Delta}' \ \ \mathbf{D}']$ is n × p design matrix with rank m (< p), p = 1+v +b, **1** is n × 1 vector of ones and $\boldsymbol{\theta} = [\mu \ \ \boldsymbol{\tau}' \ \ \boldsymbol{\beta}']$ is the parameter vector. $\mathbf{\Delta}'$ is a n × (v +1) incidence matrix of observations vs treatments, $\mathbf{D}'$ is a n × b incidence matrix of observations vs blocks, μ is the general mean, $\boldsymbol{\tau} = (\tau_0, \tau_1, ..., \tau_v)'$ is the vector of treatment effects, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_b)'$ is the vector of block effects. Suppose that the vector **y** contains the observations in such a way that the observations pertaining to the control treatment are written first followed by the observations pertaining to the test treatments. $\mathbf{\Delta}'\mathbf{1} = \mathbf{1}$, $\mathbf{\Delta}\mathbf{1} = \mathbf{r} = (r_0, r_1, ..., r_v)'$, where r_0 is the replication number of control treatment and r_i is the replication number of the i^{th} test treatment, i = 1, 2, ..., v.

$\mathbf{D'1} = \mathbf{1}$, $\mathbf{D1} = \mathbf{k} = (k_1, k_2, ..., k_b)'$ and $\mathbf{\Delta D'} = \mathbf{W}_d = ((n_{dhj}))$ is the (v+1) × b incidence matrix of the treatments vs blocks, $h = 0, 1, 2, \cdots, v$. Rewrite $\mathbf{W}_d$ as $\mathbf{W}_d = \begin{bmatrix} \mathbf{N}'_{d0} \\ \mathbf{N}_d \end{bmatrix}$, where $\mathbf{N}'_{d0}$ is a 1 × b incidence vector of the control treatment versus blocks and $\mathbf{N}_d$ is a v × b incidence matrix of test treatments vs blocks. Rewrite again the model (2.2) as

$$\mathbf{y} = [\mathbf{X}_1 \quad \mathbf{X}_2] \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} + \mathbf{e} \tag{2.3}$$

where $\mathbf{X}_1 = \mathbf{\Delta}'$ and $\mathbf{X}_2 = [\mathbf{1} \quad \mathbf{D}']$, $\boldsymbol{\theta}_1 = \boldsymbol{\tau}$ contains v+1 parameters of interest to the experimenter, and $\boldsymbol{\theta}_2 = [\mu \quad \boldsymbol{\beta}']'$ is a (1+b) vector of nuisance parameters in the model. Using the principle of ordinary least squares, the reduced normal equations for estimating the treatment effects are given by

$$\mathbf{C}_d \boldsymbol{\tau} = \mathbf{Q}_d \tag{2.4}$$

where $$\mathbf{C}_d = \mathbf{X}'_1 \mathbf{B} \mathbf{X}_1 = \mathbf{\Delta B \Delta}', \ \mathbf{Q}_d = \mathbf{X}'_1 \mathbf{B y} \tag{2.5}$$

and $\mathbf{B} = \mathbf{I} - \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{X}_2)^- \mathbf{X}'_2$ is symmetric and idempotent. We can also write

$$\mathbf{C}_d = \mathbf{R}_d - \mathbf{N}_d \mathbf{K}_d^{-1} \mathbf{N}'_d = \begin{bmatrix} e & \mathbf{a}' \\ \mathbf{a} & \mathbf{M}_d \end{bmatrix} \tag{2.6}$$

where $\mathbf{\Delta\Delta}' = \mathbf{R}_d = \text{diag}\,(r_0, r_1, ..., r_v)$, $DD' = K_d = \text{diag}(k_1, k_2, ..., k_b)$

$e = r_0 - N'_{d0} K_d^{-1} N_{d0}$, $\mathbf{a} = -\mathbf{N}_d \mathbf{K}_d^{-1} \mathbf{N}_{d0}$ and $M_d = R - N_d K_d^{-1} N'_d$

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

Following Gopalan and Dey [9], Bhar [1] and Bhar and Gupta [2] and rearranging the elements of matrix **X**, we have

$$\mathbf{X} = [\mathbf{D}' \quad \mathbf{1} \quad \mathbf{\Delta}'], \ \boldsymbol{\theta} = [\boldsymbol{\beta}' \quad \mu \quad \boldsymbol{\tau}']', \ (\mathbf{X'X}) = \begin{bmatrix} \mathbf{K}_d & \mathbf{k} & \mathbf{N}'_d \\ \mathbf{k}' & n & \mathbf{r}' \\ \mathbf{N}_d & \mathbf{r} & \mathbf{R}_d \end{bmatrix} \text{ and}$$

$$(\mathbf{X'X})^- = \begin{bmatrix} \mathbf{L}_{b\times b} & -\mathbf{0}_{b\times 1} & -\mathbf{E}_{b\times(p+1)} \\ -\mathbf{0}_{1\times b} & 0 & \mathbf{0}_{1\times(p+1)} \\ -\mathbf{E}'_{(p+1)\times b} & \mathbf{0}_{(p+1)\times 1} & \mathbf{P}_{(p+1)\times(p+1)} \end{bmatrix} \tag{2.7}$$

where $\mathbf{L} = \mathbf{K}_d^{-1} + \mathbf{K}_d^{-1}\mathbf{N}_d'\mathbf{C}_d^{-}\mathbf{N}_d\mathbf{K}_d^{-1}$, $\mathbf{E} = \mathbf{K}_d^{-1}\mathbf{N}_d'\mathbf{C}_d^{-}$ and $P = \mathbf{C}_d^{-}$.

Suppose that the u^{th} observation pertaining to the i^{th} treatment in j^{th} block, $u = 1,2,\ldots,n$; $i = 0,1,2,\ldots,v$; $j = 1,2,\ldots,b$ is denoted by $\mathbf{x}_u'\boldsymbol{\theta}$, where $\mathbf{x}_u'$ is the u^{th} row of the matrix $\mathbf{X}$. The estimated value of $\mathbf{x}_u'\boldsymbol{\theta}$ is given by $\mathbf{x}_u'\hat{\boldsymbol{\theta}} = \hat{y}_u = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j$, and using (2.7) we get

$$\operatorname{Var}\left(x_u'\hat{\theta}\right) = \sigma^2 x_u'(X'X)^{-} x_u = L_{jj} - 2E_{j0} + P_{00} \tag{2.8}$$

where L_{jj} is the j^{th} diagonal element of the matrix $\mathbf{L}$. E_{jh} and $P_{hh'}$ are respectively the elements of j^{th} row and h^{th} column of the matrices $\mathbf{E}$ and $\mathbf{P}$.

3. Detection of Outlier in Block Designs for Test Treatments-Control Comparisons

We now proceed to obtain a general form of Cook-statistic for detection of outliers in the data obtained from designed experiments pertaining to the experimental setting for comparing test treatments with a control treatment. The contrasts of interest, $\mathbf{P}\boldsymbol{\theta}_1$, are of the type $\tau_0 - \tau_i$, $\forall\ i = 1,2,\ldots,v$, where $\mathbf{P} = [\mathbf{1}_v \ \ -\mathbf{I}_v]$. The BLUE of $\mathbf{P}\boldsymbol{\theta}_1$ is $\mathbf{P}\hat{\boldsymbol{\theta}}_1$. For a connected design the dispersion matrix of $\mathbf{P}\hat{\boldsymbol{\theta}}_1$ is $D(\mathbf{P}\hat{\boldsymbol{\theta}}_1) = \sigma^2\mathbf{P}\mathbf{C}_d^{-}\mathbf{P}'$.

Suppose that the u^{th} observation is suspected to be an outlier, $u = 1,2,\cdots,n$. In the presence of an outlying observation the mean shift model can be written as

$$\mathbf{y} = \mathbf{Z}\gamma + \mathbf{e} \tag{3.1}$$

where $\mathbf{Z} = [\mathbf{X} \ \ \mathbf{u}]$ and $\mathbf{u} = [0 \ \ 0\ldots \ \ 1(u^{th})\ldots \ \ 0]'$, $\gamma = (\theta' \ \ \delta)'$ and δ is a non-zero scalar quantity.

Following Bhar [1] the Cook statistic for these experimental situations is defined as

$$D_1 = \frac{(\mathbf{P}\hat{\boldsymbol{\theta}}_1 - \mathbf{P}\hat{\boldsymbol{\theta}}_{1(u)})'[D(\mathbf{P}\hat{\boldsymbol{\theta}}_1)]^{-}(\mathbf{P}\hat{\boldsymbol{\theta}}_1 - \mathbf{P}\hat{\boldsymbol{\theta}}_{1(u)})}{\operatorname{Rank}[D(\mathbf{P}\hat{\boldsymbol{\theta}}_1)]}$$

where $\mathbf{P}\hat{\boldsymbol{\theta}}_{1(u)}$ is the BLUE of $\mathbf{P}\boldsymbol{\theta}_1$ obtained after eliminating the u[th] outlying observation. We then have the following result

Result 1.{Lemma 3.4: Bhar [1]}

$$\mathbf{P}(\hat{\boldsymbol{\theta}}_1 - \hat{\boldsymbol{\theta}}_{1(u)}) = \mathbf{P}\mathbf{C}_d^{-}\mathbf{X}_1'\mathbf{B}\mathbf{u}(\mathbf{u}'\mathbf{V}\mathbf{u})^{-1}\mathbf{u}'\mathbf{V}\mathbf{y}$$

where $\mathbf{V} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{B} - \mathbf{S}$; $\mathbf{S} = \mathbf{B}\mathbf{X}_1\mathbf{C}_d^{-}\mathbf{X}_1'\mathbf{B}$

The matrices $\mathbf{V}$ and $\mathbf{S}$ are symmetric and idempotent. Therefore

$$D_1 = \frac{\mathbf{u}'\mathbf{B}\mathbf{X}_1\mathbf{C}_d^{-}\mathbf{P}'(\mathbf{P}\mathbf{C}_d^{+}\mathbf{P}')^{-}\mathbf{P}\mathbf{C}_d^{-}\mathbf{X}_1'\mathbf{B}\mathbf{u}\hat{\delta}^2}{v\sigma^2}$$

where $\hat{\delta}$ is the estimated value of δ and is given by $\hat{\delta} = (\mathbf{u}'\mathbf{V}\mathbf{u})^{-1}\mathbf{u}'\mathbf{V}\mathbf{y}$

A g-inverse of $\mathbf{C}_d$ as given in (2.6) is $\mathbf{C}_d^{-} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_d^{-1} \end{bmatrix}$ and, therefore, $\mathbf{P}\mathbf{C}_d^{-}\mathbf{P}' = \mathbf{M}_d^{-1}$, where $\mathbf{M}_d^{-1}$ is the information matrix corresponding to the test treatments only.

Therefore, $$D_1 = \frac{\mathbf{u}'\mathbf{B}\mathbf{X}_1\mathbf{C}_d^{-}\mathbf{P}'\mathbf{M}_d\mathbf{P}\mathbf{C}_d^{-}\mathbf{X}_1'\mathbf{B}\mathbf{u}\hat{\delta}^2}{v\sigma^2}$$

$$= \frac{\mathbf{u}'\mathbf{T}\mathbf{u}\hat{\delta}^2}{v\sigma^2} \qquad (3.2)$$

where $\mathbf{T} = \mathbf{B}\mathbf{X}_1\mathbf{C}_d^{-}\mathbf{P}'\mathbf{M}_d\mathbf{P}\mathbf{C}_d^{-}\mathbf{X}_1'\mathbf{B}$ (3.3)

$$\text{Now } \mathbf{C}_d^{-}\mathbf{P}' = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_d^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{1}_v' \\ -\mathbf{I}_v \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{0} \\ \mathbf{M}_d^{-1} \end{bmatrix}$$

and $\mathbf{C}_d^{-}\mathbf{P}'\mathbf{M}_d\mathbf{P}\mathbf{C}_d^{-} = \mathbf{C}_d^{-}$

Therefore, from (3.3) we have

$$\mathbf{T} = \mathbf{B}\mathbf{X}_1\mathbf{C}_d^{-}\mathbf{X}_1'\mathbf{B} = \mathbf{S} \qquad (3.4)$$

Using (3.4) in (3.2) and replacing σ^2 by $\hat{\sigma}^2$, we get

$$D_1 = \frac{\mathbf{u}'\mathbf{S}\mathbf{u}\hat{\delta}^2}{v\hat{\sigma}^2} = \frac{s_{11}\hat{\delta}^2}{v\hat{\sigma}^2} \qquad (3.5)$$

which is the same as obtained by Bhar and Gupta [2] for a general block design. Thus the general expression of Cook-statistic for detection of a single outlier in block designs for comparing test treatments with a single control treatment remains the same as that of general block designs for making all the possible paired comparisons. The difference, however, is that in designs for making test treatments-control treatment comparisons the outlier may be an observation pertaining to a test treatment or the control treatment. Through some algebraic simplifications, it can be shown easily that the Cook-statistic for any i^{th} observation ($i = 1,2,\ldots, n$) from the j^{th} block is given by

$$D_i = \frac{\mathbf{u}_0'\mathbf{C}_d^-\mathbf{u}_0}{\left(1-\mathbf{u}_0'\mathbf{C}_d^-\mathbf{u}_0\right)} \cdot \frac{t_i^2}{v}$$

$$= \frac{s_{ii}}{\dfrac{k_{j(i)}-1}{k_{j(i)}} - s_{ii}} \cdot \frac{t_i^2}{v} \qquad (3.6)$$

where s_{ii} is the i^{th} diagonal element of matrix $\mathbf{S}$ and is given by

$$s_{ii} = \{(k_{j(i)}-1)/k_{j(i)}\}\mathbf{u}_0'\mathbf{C}_d^-\mathbf{u}_0 \,;\; \mathbf{u}_0\{k_{j(i)}(k_{j(i)}-1)\}^{1/2} = \frac{1}{k_{j(i)}}\begin{bmatrix} k_{j(i)}-1 \\ -\mathbf{f} \end{bmatrix}$$

where $k_{j(i)}$ is the j^{th} block which contains the i^{th} observation suspected to be an outlier, $\mathbf{f}$ is a $v \times 1$ vector of incidence of v treatments in the j^{th} block, t_i is the i^{th} studentized residual.

4. *Robustness Aspects of Balanced Treatment Incomplete Block Designs*

For studying the robustness of designs for comparing several test treatments with a control treatment against the presence of a single outlier, we may use average Cook statistic as used by Bhar and Gupta [2]. However, for studying the problem of outlier in designs for making test treatments vs control treatment comparison, we cannot use the Cook statistic averaged over all the n observations. If the outlying observation pertains to the control treatment then we take the average of Cook statistic over all the n_0 observations pertaining to control treatment; on the other hand, if the outlying observation arises from the test treatments then we take the average value of Cook statistic over all possible $(n - n_0)$ observations.

Using (3.6), the average value of Cook-statistic when outlying observation arises from control treatment is

$$\overline{D}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} \frac{s_{ii}}{\frac{k_{j(i)}-1}{k_{j(i)}} - s_{ii}} \cdot \frac{t_i^2}{v} \tag{4.1}$$

Similarly, the average value of Cook-statistic when outlying observation comes from test treatments is

$$\overline{D}_t = \frac{1}{(n-n_0)}\sum_{i=n_0+1}^{n} \frac{s_{ii}}{\frac{k_{j(i)}-1}{k_{j(i)}} - s_{ii}} \cdot \frac{t_i^2}{v} \tag{4.2}$$

It is seen that both the averages in (4.1) and (4.2) are weighted sum of squares of t_i's. Write the matrix **S** in partitioned form as $\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}$ where $\mathbf{S}_{11}$ is the sub-matrix corresponding to n_0 observations on control treatment and $\mathbf{S}_{22}$ is the sub matrix corresponding to $n - n_0$ observations on test treatments.

We now describe the robustness criterion as in case of general block designs as follows: A design is robust against the presence of a single outlier pertaining to control treatment if all the diagonal elements of the sub matrix $\mathbf{S}_{11}$ are equal; similarly, a design is robust against the presence of a single outlier pertaining to a test treatment if all the diagonal elements of the sub matrix $\mathbf{S}_{22}$ are equal.

Now depending upon whether the outlying observation pertains to a control treatment or a test treatment, (2.8) may be simplified as below

Case 1. When the outlying observation pertains to the control treatment and appears in the j^{th} block

For this case the expression in (2.8) is $\text{Var}(\mathbf{x}'_u\hat{\boldsymbol{\theta}}) = L_{jj} - 2E_{j0} + P_{00}$, where E_{j0} is the element in the j^{th} row and first column of the matrix **E** and P_{00} is the first diagonal element of the matrix **P**, j = 1,2, ..,b.

Case 2. When the outlying observation pertains to the test treatment and appears in the j^{th} block

For this case the expression in (2.8) is $\text{Var}(\mathbf{x}'_u\hat{\boldsymbol{\theta}}) = L_{jj} - 2E_{jh} + P_{hh}$, h = 1, 2, ..., v; j = 1, 2, ..., b, where E_{jh} is the element in the j^{th} row and h^{th} column of the matrix **E** and P_{hh} is the h^{th} diagonal element of the matrix **P**.

It appears difficult to establish the equality of variance of $\mathbf{x}'_u\hat{\boldsymbol{\theta}}$, $\forall$ u =1,2,..., n for a general block design set up. Therefore, we restrict our study to only those block designs that are variance balanced for estimation of test treatments–control treatment contrasts. A block design is said to be variance balanced for the estimation of test treatments vs control treatment contrasts if it permits the estimation of these contrasts with the same variance and covariance between any two estimated test treatments vs control treatment contrasts is also same. In a proper block design set up, these designs have been termed as balanced treatment incomplete block (BTIB) designs. These are defined as below

Definition 4.1. An arrangement of v test treatments and a control treatment in b blocks each of size $k < (v + 1)$ is said to be a BTIB design if

(i) $\sum_{j=1}^{b} n_{ij} n_{i'j} = \lambda$, a constant $\forall\ i \neq i' = 1, 2, \ldots, v$

(ii) $\sum_{j=1}^{b} n_{0j} n_{ij} = \lambda_0$, a constant $\forall\ i = 1, 2, \ldots, v$

For a BTIB design, the $\mathbf{C}_d$ matrix given in (2.6) is

$$\mathbf{C}_d = \begin{bmatrix} \frac{v\lambda_0}{k} & -\frac{\lambda_0}{k}\mathbf{1}'_v \\ -\frac{\lambda_0}{k}\mathbf{1}_v & \frac{v\lambda + \lambda_0}{k}\mathbf{I} - \frac{\lambda}{k}\mathbf{1}\mathbf{1}' \end{bmatrix} \tag{4.3}$$

with $\mathbf{M}_d = \frac{v\lambda + \lambda_0}{k}\mathbf{I} - \frac{\lambda}{k}\mathbf{1}\mathbf{1}'$. It is easy to see that a generalized inverse of $\mathbf{C}_d$, as given in (2.6) is $\mathbf{C}_d^- = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \mathbf{M}_d^{-1} \end{bmatrix}$

It is easily seen that $\mathbf{M}_d^{-1} = \frac{k}{(\lambda_0 + v\lambda)}\mathbf{I} + \frac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)}\mathbf{1}\mathbf{1}'$

We now have $P_{00} = 0$, $P_{ii} = i^{th}$ diagonal element of $\mathbf{M}_d^{-1} = \frac{k}{(\lambda_0 + v\lambda)} + \frac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)}$

Also $\mathbf{E} = \mathbf{K}_d^{-1}\mathbf{N}'_d\mathbf{C}_d^- = \mathbf{K}_d^{-1}[\mathbf{N}_{d0} \quad \mathbf{N}'] \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_d^{-1} \end{bmatrix} = [\mathbf{0} \quad \mathbf{K}_d^{-1}\mathbf{N}'\mathbf{M}_d^{-1}]$

Therefore, $E_{j0} = 0$ and

$$E_{jh} = \frac{1}{(\lambda_0 + v\lambda)} + \frac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)}, \text{ for all } h = 1, 2, \ldots, v$$

$$\mathbf{L} = \mathbf{K}_d^{-1} + \mathbf{K}_d^{-1}\mathbf{N}_d'\mathbf{C}_d^{-}\mathbf{N}_d\mathbf{K}_d^{-1}$$
$$= \mathbf{K}_d^{-1} + \mathbf{K}_d^{-1}\mathbf{N}'\mathbf{M}_d^{-1}\mathbf{N}\mathbf{K}_d^{-1}$$

Therefore, $L_{jj} = \dfrac{1}{k} + \dfrac{1}{k(\lambda_0 + v\lambda)}\sum_{i=1}^{v} n_{ji}^2 + \dfrac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)}$

Let us assume that the design is binary with respect to test treatments, then

$\sum_{i=1}^{v} n_{ji}^2 = \sum_{i=1}^{v} n_{ij} = k$ and so

$$L_{jj} = \frac{1}{k} + \frac{1}{(\lambda_0 + v\lambda)} + \frac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)}$$

Thus $\dfrac{1}{\sigma^2}\text{Var}(\hat{y}_u) = \dfrac{1}{k} + \dfrac{1}{(\lambda_0 + v\lambda)} + \dfrac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)} \quad \forall\, u = 1, 2, \ldots, n_0$

and $\dfrac{1}{\sigma^2}\text{Var}(\hat{y}_u) = \dfrac{1}{k} + \dfrac{1}{(\lambda_0 + v\lambda)} + \dfrac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)} - \dfrac{2}{(\lambda_0 + v\lambda)} - \dfrac{2k\lambda}{\lambda_0(\lambda_0 + v\lambda)}$

$$+ \frac{k}{(\lambda_0 + v\lambda)} + \frac{k\lambda}{\lambda_0(\lambda_0 + v\lambda)}$$

$$= \frac{1}{k} - \frac{1}{(\lambda_0 + v\lambda)} + \frac{k}{(\lambda_0 + v\lambda)} \quad \forall\, u = n_0 + 1, n_0 + 2, \ldots, n$$

which is constant for given v, k, λ, λ_0. We, therefore, have the following result

Theorem 4.1. All BTIB designs that are binary with respect to test treatments are robust against the presence of a single outlier.

In the present investigation, we have made an attempt to develop test statistic for detection of outlier(s) in block designs for making test treatments-control treatment comparisons and identified robust designs against the presence of a single outlier. A design is said to be robust if it minimizes the average value of Cook statistic. It may be noted that the criteria of robustness is dependent on design matrix **X** alone and doesn't involve the observation vector **y**. As a result the value of F-statistic for studying the significance of treatment effects may get affected by the presence of outlier(s) even if the design is robust according to the above criterion. Therefore, besides detection of outlier(s) and identification of robust designs in presence of outlier(s), it is essential to develop some estimation/analytical procedures so that inference on the parameters of interest does not change. One way to deal with such situations is to develop robust procedure of estimation of treatment contrasts. However, this is beyond the scope of the present investigation. In the absence of robust estimation procedures, one may think of either deleting the observation(s) identified as outlier(s) or carrying out the analysis of covariance by defining a covariate for each of the outlying observation. Bhar and Gupta [2] have shown that the reduced normal equations under the analysis of covariance model or with suspected observation(s) deleted from the model are same. Therefore, both the above alternatives are same and one may use either of them. As the reduced normal equations under the analysis of covariance model or with suspected observation(s) deleted from the model are same, therefore, for a design to be robust against the presence of a single outlier, it must remain connected after one missing observation. The results on robustness of block designs for making test treatments-control treatment comparisons against a missing observation obtained by Srivastava *et al.* [14] may be quite useful in these situations. Keeping in view the above, we can say that all BTIB designs, binary with respect to test treatments and robust against a single missing observation, are robust against the presence of a single outlier as per the criterion of minimization of average Cook-statistic.

ACKNOWLEDGEMENTS

The authors wish to thank the referee for many useful suggestions that helped to improve an earlier version of this paper.

REFERENCES

[1] Bhar, L. (1997). *Outliers in Experimental Designs*. Ph.D. Thesis, I.A.R.I., New Delhi.

[2] Bhar, Lalmohan and Gupta, V.K. (2001). A useful statistic for studying outliers in experimental designs. *Sankhyā*, **B63**, 338-350.

[3] Box, G.E.P. and Draper, N.R. (1975). Robust designs. *Biometrika*, **62(2)**, 347-352.

[4] Cook, D.R. (1977). Detection of influential observation in linear regression. *Technometrics*, **19(1)**, 15-18.

[5] Dey, A. (1986). *Theory of Block Designs*. Wiley Eastern Limited, New Delhi.

[6] Ghosh, S. (1983). Influential observations in view of design and inference. *Commu. Stat.-Theory and Methods*, **12(14)**, 1675-1683.

[7] Ghosh, S. (1989). On two methods of identifying influential sets of observations. *Statist. Prob. Letters*, **7**, 241-245.

[8] Ghosh, S. and Kipngeno, Wilson, A.K. (1985). On the robustness of optimum balanced 2^m factorial designs of resolution V (given by Srivastava and Chopra) in the presence of outliers. *J. Statist. Plann. Inf.*, **11**, 119-129.

[9] Gopalan, R. and Dey, A. (1976). On robust experimental designs. *Sankhyā*, **B38,** 297 -299.

[10] Gupta, V.K. and Parsad, R. (2001). Block designs for comparing test treatments with control treatments- an overview. *Statistics and Applications*, **3,** 133-146.

[11] Hedayat, A.S., Jacroux, M. and Majumdar, D. (1988). Optimal designs for comparing test-treatments with a control. *Statist. Sc.*, **3,** 462-491.

[12] Majumdar, D. (1996). Optimal and efficient treatment-control designs. In : *Handbook of Statistics,* **13** (Ed. S. Ghosh and C.R. Rao), Elsevier Science B.V., 1007-1053.

[13] Singh, G., Gupta, S. and Singh, M. (1987). Robustness of row-column designs. *Statist. Prob. Letters*, **5,** 421-424.

[14] Srivastava, R., Parsad, R. and Gupta, V.K. (1996). Robustness of block designs for making test treatments-control comparisons against a missing observation. *Sankhyā*, **B58(3)**, 407-413.