BALANCED INCOMPLETE BLOCK DESIGN – A REVIEW AND ITS ANALYSIS IN COMPLETE DATA AND WITH ONE MISSING OBSERVATION

A.K. Kaushik

Department of Statistics, J.V. College, Baraut (U.P.) India
E-mail: ashok_kshk@yahoo.com

Abstract: When the size of the experimental material is not sufficient to accommodate all the treatments, we require incomplete block designs to test various treatments under study in agricultural and biological sciences. The author of the present paper has discussed one of the incomplete block design, namely, Balanced Incomplete Block Design (BIBD). He has tried to present the review of the available literature on BIBD in brief, its analysis in case of complete data, and in case of one missing observation as well. The subject matter discussed here is not entirely new, but its presentation is new. However, the method for the analysis of BIBD in presence of one missing observation has been developed by him in 1992 in his unpublished Ph. D thesis. The Complex mathematical expressions are avoided in the present paper, and only simple expressions are provided to analyze the data. The methods are also supported by suitable examples. This will be of great help to the investigators engaged in agriculture and biological sciences.

Keywords: Adjusted treatment total, Adjusted treatment mean, Bias, Varietal trials

INTRODUCTION

Experimentation and making inferences are twin essential features of general scientific methodology. Statistics as a scientific discipline is mainly designed to achieve these objectives. The methodology for making inferences has three main aspects.

1. It derives methods for drawing inferences from observations when these are not exact but subject to variation.
2. It specifies methods for collection of data appropriately so that the assumptions for the application of appropriate statistical methods to them are satisfied.
3. The techniques for proper interpretation of results are devised.

A good coverage of these is available in Fisher (1953), Giri (1976), and Scheffe (1959). Mainly three types of experiments require statistical designing. These are (i) factorial experiments, (ii) varietal trials, and (iii) bio-assays. *Varietal trials* are primarily agricultural experiments to select a few varieties of a crop through experimentation over a number of varieties which are better than the rest in respect of some economic character. These experiments are generally conducted in complete block designs like RBD or LSD. These designs are used only when the number of treatments is 12 or less. We also know that the precision of the estimate of a treatment effect depends on the number of replications of the treatment. That is, larger is the number of replications, the more is the precision. A similar thing holds for the precision of the estimate of the difference between two treatment effects. This consideration has been exploited to construct designs for varietal or similar trials with large numbers of treatments so as to reduce the block size. Here comes the concept of incomplete block designs. *A block is said to be incomplete in a design if the number of plots in the block is less than the number of treatments*. In order to ensure equal or nearly equal precision of the comparisons of different pairs of treatments, the treatments are so allotted to the different blocks that each pair of the treatments has the same or nearly the same number of replications and each treatment has an equal number of replications. When the number of replications of all pairs of treatments in a design is same, then an important series of designs known as *balanced incomplete block design* (B.I.B.D) is obtained.

An incomplete block design with v treatments distributed over b blocks, each of size k (k < v), such that each treatment occurs in r blocks, no treatment occurs more than once in a block, and each pair of the treatments occurs together in λ blocks, is called a balanced incomplete block design (B.I.B.D). The symbols v, b, r, k, and λ are called parameters of the design. These parameters satisfy the following relations

(i) $v r = b k$, (ii) $\lambda (v - 1) = r (k - 1)$, (iii) $b \geq v$.

If b = v and r = k then the BIBD is called *symmetric BIBD.* It is seen that the blocks of the designs

$v = s^2$, $b = s (s + 1)$, $r = s + 1$, $k = s$, $\lambda = 1$

where s is a prime or power of a prime, can be divided into (s + 1) groups of s blocks each, such that in each group each of the treatments is replicated once. Such types of designs are called *resolvable.* Again, if the number of treatments common between any two blocks is belonging to two different groups of a resolvable design is constant, then such designs are called *affine resolvable designs.*

For resolvable designs no solution exist if
$b < v – r + 1$.

Fisher's inequality has established that the solutions of designs with b < v are not possible. No set of parameters of a BIBD with b divisible by r is possible where $b < v – r + 1$.

BIBD was first devised by Yates (1936). Later on Fisher, Yates and Bose (1939) jointly solved its construction problems. It was found that BIBD were not always suitable for varietal trials because these designs requires large number of replications and further, suitable designs are not available for all number of treatments. To overcome such difficulties, Yates (1936) evolved another series of incomplete block designs which he called *lattice designs.* Bose and Nair (1939) evolved some another type of incomplete block designs which they called *partially incomplete block designs.* Later on, some other incomplete block designs namely *re-inforced incomplete block designs, circular designs* were obtained by Das (1958) and Giri (1958). The discussion on these series of designs is beyond the scope of present study.

It is important to note that BIBD is used when the experimental material is not of sufficient size to accommodate all the treatments in the blocks.

MATERIAL AND METHODS

The present discussion does not require the detailed mathematical version of the analysis of BIBD. Here, the author presents the list of expressions used in the analysis of numerical data in a simple and lucid manner so that any researcher can understand and use them easily in case of complete and incomplete data.

Section 1: Analysis of BIBD in case of Complete data:

With the definition of BIBD, as given above, the appropriate model is

$$Y_{ij} = \mu + t_i + b_j + e_{ij} \qquad \text{…. (2.1.1)}$$

i = 1, 2, …., v ; j = 1, 2, …., r ;

with usual notations. For complete set of data, we have following expressions.

$$\text{CF} = \frac{G^2}{n}$$

$$\text{Total Sum of Square} = \sum_{i=1}^{v}\sum_{j=1}^{b} Y_{ij}^2 - CF$$

with (v r – 1) df only.

$$\text{Block Sum of Square (unadjusted)} = \sum_{j=1}^{b}\frac{B_j^2}{k} - CF ,$$

with (b – 1) df only.

$$\text{Treatment Sum of Square} = \frac{k}{\lambda v}\sum_{i=1}^{v} Q_i^2$$

with (v – 1) df only.

$$\text{Where } Q_i = T_i - \frac{1}{k}\sum_{j=1}^{b} n_{ij} B_j$$

=Adjusted treatment totals.

Error Sum of Square = Total Sum of Square – Block Sum of Square – Treatment Sum of Square with (v r – b – v + 1) df only.

Section 2: Analysis of BIBD in presence of One Missing Observation:

Without any loss of generality, we may assume that the missing observation belongs to 1st treatment in 1st block. Also, that the first k-treatments are allotted to the 1st block. The appropriate model for the analysis of such data will be

$$Y_{ij} = \mu + t_i + b_j + e_{ij} \qquad \text{…. (2.2.1)}$$

i = 1, 2, …., v ; j = 1, 2, …., r ;

with usual notations. For incomplete set of data, we have to estimate the missing observation by using the following expressions :

$$\hat{Y}_1 = \frac{\lambda v B_1 + k(kQ_1 - Q_1')}{(k-1)(\lambda v - k)} \quad \ldots (2.2.2)$$

where $Q_1' = Q_{1(Y)} + Q_{2(Y)} + \ldots.. + Q_{k(Y)}$ = Sum of all the adjusted treatment totals of the treatments falling in the 1st block.

Error Sum of Square = E. S. S =

$$\left(\hat{Y}_1^2 + \sum_a Y_a^2\right) - \frac{1}{k}\left((B_1 + \hat{Y}_1)^2 + \sum_{j=2}^{b} B_j^2\right) - \frac{k}{\lambda v}\sum_{i=1}^{v} Q_i^2 \quad \ldots (2.2.3)$$

With (n – b – v) df.

Under $H_0 : t_1 = t_2 = \ldots\ldots = t_v$, i.e. the treatments are homogeneous, the model (2.2.1) reduces to

$$Y_{ij} = \mu + b_j + e_{ij} \quad \ldots. (2.2.4)$$

The new estimate of the missing value under (2.2.2) will be

$$\overset{*}{Y}_1 = \frac{B_1}{(k-1)} \quad \ldots (2.2.5)$$

The new error sum of square under (2.2.2) will be

$$E_0.\ S.S. = \left(\overset{*}{Y}{}_1^2 + \sum_a Y_a^2\right) - \frac{1}{k}\left((B_1 + \overset{*}{Y}_1)^2 + \sum_{j=2}^{b} B_j^2\right) - \frac{k}{\lambda v}\sum_{i=1}^{v} Q_i^2 \quad \ldots (2.2.6)$$

With (n – b – 1) df.

Treatment Sum of Square=

$$E_0.\ S.\ S. - E.\ S.\ S. = \frac{k}{\lambda v}\sum_{i=1}^{v} Q_i^2 - \text{Bias}, \quad \ldots. (2.2.7)$$

with (v – 1) df only.

$$\text{Bias} = \frac{(k-1)}{k}(\hat{Y}_1 - \overset{*}{Y}_1)^2 = \frac{(k-1)}{k}\left(\hat{Y}_1 - \frac{B_1}{(k-1)}\right)^2 \quad \ldots (2.2.8)$$

$$\underset{u=2,3,\ldots,k}{v(\hat{t}_i - \hat{t}_u)} = \frac{2k\sigma^2}{\lambda v} + \frac{k^3\sigma^2}{\lambda v(k-1)(\lambda v - k)} \quad \ldots (2.2.9)$$

$$\underset{w=k+1,k+2,\ldots,v}{v(\hat{t}_i - \hat{t}_u)} = \frac{2k\sigma^2}{\lambda v} + \frac{k(k-1)\sigma^2}{\lambda v(\lambda v - k)} \quad \ldots (2.2.10)$$

$$v(\hat{t}_u - \hat{t}_w) = \frac{2k\sigma^2}{\lambda v} + \frac{k\sigma^2}{\lambda v(k-1)(\lambda v - k)} \quad \ldots (2.2.11)$$

$$\underset{u\neq u'}{v(\hat{t}_i - \hat{t}_u)} = \underset{w\neq w'}{v(\hat{t}_i - \hat{t}_u)} = \frac{2k\sigma^2}{\lambda v} \quad \ldots (2.2.12)$$

$$\text{Relative Efficiency} = \frac{(v-1)(\lambda v - k)}{\{(v-1)(\lambda v - k) + k\}} \quad \ldots (2.2.13)$$

Relative Loss in Efficiency=

$$\frac{k}{\{(v-1)(\lambda v - k) + k\}} \quad \ldots (2.2.14)$$

Illustration 1: Following table gives the results of an experiment conducted in a BIBD for comparing 7 treatments in 7 blocks of 3 units each.

Table 1:

Treatment	Blocks						
	1	2	3	4	5	6	7
1	50	42	91	–	–	–	–
2	–	–	118	94	94	–	–
3	76	–	–	64	–	80	–
4	–	–	72	–	–	53	31
5	44	–	–	–	65	–	54
6	–	102	–	–	119	92	–
7	–	38	–	38	–	–	37

The above shown design is a symmetric BIBD with parameters (v = 7 = b, r = 7 = k, λ = 1).

H_0 : Treatments are homogeneous.

The results are analyzed in the following table :

Sl. No.	T_i	Blocks	B_j	$\sum n_{ij} B_j$	Q_i	Q_i^2	T_i^2	B_j^2
1	183	1,2,3	170	633	– 28.00	781.00	33489	28900
2	306	3,4,5	182	755	54.33	2951.75	93636	33124
3	220	1,4,6	281	591	21.00	529.00	48400	78961
4	156	3,6,7	196	628	– 53.33	2844.09	24336	38416
5	163	1,5,7	278	570	– 27.00	729.00	26569	77284
6	313	2,5,6	225	685	84.67	7169.00	97969	50625
7	113	2,4,7	122	500	– 53.67	2880.47	12769	14844
Total	**1454**		**1454**	**4362**	**0**	**17887.31**	**337168**	**322194**

C.F.= $\frac{G^2}{n} = \frac{(1454)^2}{21}$ = 100672.19, Total Sum of Square = $\sum_{i=1}^{v}\sum_{j=1}^{b} Y_{ij}^2 - CF$ = 15057.81,

Block Sum of Square (unadjusted) = $\sum_{j=1}^{b} \frac{B_j^2}{k} - CF$ = 6725.81,

Treatment Sum of Square = $\frac{k}{\lambda v}\sum_{i=1}^{v} Q_i^2$ = 7665.99

E. S. S. = 15057.81 – 6725.81 – 7665.99 = 666.01

ANOVA Table

Source of Variation	d.f.	S. S.	M. S. S.	Variance Ratio	$F_{tab., 0.05}$
Treatments (adjusted)	6	7665.99	1277.67	15.347**	3.58
Blocks (unadjusted)	6	6725.81	1120.97	13.46	-
Error	8	666.01	83.25	-	-
Total	**20**	**15057.80**	**-**	**-**	**-**

It is clear that F_{cal} = 15.347 > 3.58 , we conclude that null hypothesis H_0 is rejected at 5 % level of significance i.e. the treatments differ significantly.

Adjusted treatment means= $\bar{y}_{..} + \frac{k}{\lambda v}Q_i = \frac{G}{bk} + \frac{k}{\lambda v}Q_i = 69.24 + \frac{3}{7}Q_i$

Which gives treatment means as

$\bar{T}_1$ = 57.24, $\bar{T}_2$ = 92.53, $\bar{T}_3$ = 79.10, $\bar{T}_4$ = 46.38, $\bar{T}_5$ = 57.67, $\bar{T}_6$ = 105.52, $\bar{T}_7$ = 46.24

It is also clear that 6th treatment is the best treatment followed by 2nd , 3rd , 5th , 1st , 4th, and 7th respectively. Other discussion is beyond the scope of the present study.

Illustration 2 : As an illustration for one missing observation in BIBD, I analyze the following B. I. B. Design with parameters v = 6, b = 10, r = 5, k = 3, λ = 2 with one missing observation. The missing observation belongs to 5th treatment in 9th block shown within the rectangle.

Blocks	Contents	Blocks	Contents
I	2,5,1	VI	5,6,3
II	2,3,6	VII	5,6,4
III	2,3,4	VIII	6,1,2
IV	3,4,1	IX	1,3,5
V	4,5,2	X	1,4,6

The estimate of the missing observation will be obtained by

$$\hat{Y} = \frac{\lambda v B_9 + k\{kQ_5 - Q_5'\}}{(k-1)(\lambda v - k)} = \frac{12B_9 + 3\{Q_5 - Q_5'\}}{18}$$

where B_9 = Total of all the known observations of the 9th block,

Q_5 = Adjusted treatment total of the 5th treatment,

$Q_5' = Q_{1(Y)} + Q_{3(Y)} + Q_{5(Y)}$

The error sum of square will be

$$\text{E.S.S.} = \left(\hat{Y}^2 + \sum_a Y^2\right) - \frac{1}{3}\left\{\left(B_9 + \hat{Y}\right)^2 + \sum_{j=1, j\neq 9}^{10} B_j^2\right\} - \frac{1}{4}\sum_{i=1}^{6} Q_i^2 \qquad \text{with 14 d.f. only.}$$

Under H_0 : Treatments are homogeneous. The new estimate of the missing observation and error sum of square will be

$$\overset{*}{Y} = \frac{B_9}{(k-1)} = \frac{B_9}{2}$$

$$E_0.\ \text{S.S.} = \left(\overset{*}{Y}{}^2 + \sum_a Y^2\right) - \frac{1}{3}\left\{\left(B_9 + \overset{*}{Y}\right)^2 + \sum_{j=1, j\neq 9}^{10} B_j^2\right\} \qquad \text{with 19 d.f. only}$$

Treatment S. S. = E_0. S. S. – E. S. S. = $\frac{1}{4}\sum_{i=1}^{6} Q_i^2$ – Bias with 5 d.f. only.

$$\text{Bias} = \frac{2}{3}\left(\hat{Y} - \overset{*}{Y}\right)^2 = \frac{2}{3}\left(\hat{Y} - \frac{B_9}{2}\right)^2$$

$$\underset{u=1,3}{v(\hat{t}_5 - \hat{t}_u)} = \frac{\sigma^2}{2} + \frac{\sigma^2}{8} \qquad \underset{w=2,4,6}{v(\hat{t}_5 - \hat{t}_w)} = \frac{\sigma^2}{2} + \frac{\sigma^2}{18}$$

$$\underset{u\neq u'}{v(\hat{t}_u - \hat{t}_{u'})} = \underset{w\neq w'}{v(\hat{t}_w - \hat{t}_{w'})} = \frac{\sigma^2}{2} \qquad v(\hat{t}_u - \hat{t}_w) = \frac{\sigma^2}{2} + \frac{\sigma^2}{72}$$

Relative Efficiency = $\frac{15}{16}$ Relative Loss in Efficiency = $\frac{1}{16}$

REFERENCES

Bartlett, M.S. (1937). Some examples of statistical methods of research in agriculture and applied biology. *J. Roral. Soc.*, **B4**,137-183.

Biometrics (1957). Special issue on analysis of covariance, Vol. **13** No 3.

Das, M.N. (1955). Missing plots in partially balanced and other incomplete block designs. *J. Ind. Soc. Agri. Statist*, **7**, 111-126.

Kaushik. A.K. (1992). Analysis os balanced designs in presence of several missing observations. Unpublished Ph. D. thesis submitted to Meerut University, Meerut.

Kaushik. A.K. (2010). Analysis of balanced incomplete block design in presence of one missing observation. *Int. J. Agricult. Stat. Sci.*, Vol **6**, No. **2**, pp 615-621.

Kshirsagar, A.M. (1971). Bias due to missing plots., *The American Statistician*, **25**, 47-50.

Kshirsagar, A.M. and Mckee Bonnie (1982). A unified theory of missing plots in experimental designs., *METRON*, Vol. XL-N, 3-4.