ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.6, No.7, 2016

On The Estimation of Two Missing Values in Randomized Complete Block Designs

EFFANGA, EFFANGA OKON[1] AND BASSEY, E. E.[2]

[1]DEPARTMENT OF STATISTICS, UNIVERSITY OF CALABAR, CALABAR. NIGERIA.

[2]DEPARTMENT OF STATISTICS, UNIVERSITY OF CALABAR, CALABAR. NIGERIA.

ABSTRACT:

This paper reviews the work of Bhad and Ahmed (2012), identify flaws in their model formulation and make corrections accordingly. Furthermore, the formular for the estimation of two missing values in randomized complete block designs is derived and is applied to the example given in Bhad and Ahmed (2012). The results obtained from our formular produces a better estimate of missing values than that of Bhad and Ahmed (2012).

Key words: Randomized Complete Block Design, Sum of Squares of Errors, Missing Values, Estimation, and Sources of Variability.

1. Introduction

Occasionally one or more values are missing in a randomized complete block design (RCBD) due to some reasons. Such missing values would have to be estimated before performing ANOVA because of loss of orthogonality. Researches on the methodology for the estimation of missing values abound in literature. Formular for the estimation of a single missing value in RCBD originated from the work of Yate (1933). Montgomery (1976) used the formular for single missing value iteratively to estimate two or more missing values. Other methods for estimating missing values in RCBD can be found in Dempster & et al(1977), Jarret (1978) and Murray (1986).

Bhad and Ahmed (2012) developed a mathematical programming model to estimate k missing values in the design with several sources of variation and illustrated their model with RCBD with two missing values. Their paper contains some flaws in the functional constraints and thus need to be corrected.

Furthermore, the iterative procedure for estimating two or more missing values may take several iterations to converge and thus require another method of achieving the same or approximate result in a very short time and with less computational effort.

In this paper, we correct flaws in Bhad and Ahmed (2012) and then derived a formular for the estimation of two missing values in RCBD.

2. Review of Bhad and Ahmed (2012)

The general mathematical programming model introduced in Bhad and Ahmed (2012) is as follows:

M1: $$\text{Minimize } \{f(x_i) = SSE\} \quad (1)$$

Subject to:

$$\text{Variance}(S_i) \le \sigma_i^2 \quad (2)$$

$$x_i \ge 0 \quad (3)$$

Where S_i is the i[th] source of variation and σ_i^2 is the variance of the i[th] source of variation without considering the missing values.

The model M1 was applied to randomized complete block designs (i.e. design having two sources of variations, block and treatment). Their model which considered k missing values is given as follows:

M2: $$\text{Minimize } \{SSE = x_i^2\left\{1 - \frac{1}{t} - \frac{1}{r} + \frac{1}{rt}\right\} - 2x_i\left\{\frac{Y_{.j}}{t} + \frac{Y_{i.}}{r} - \frac{Y_{..}}{rt}\right\} \quad (4)$$

Subject to:

$$\frac{1}{r}\left\{\sum(Y_{i.})^2 + x_i^2\right\} - [\frac{1}{r}(\sum Y_{i.} + x_i)]^2 \le \sigma_1^2 \quad (5)$$

$$\frac{1}{t}\left\{\sum(Y_{.j})^2 + x_i^2\right\} - [\frac{1}{t}(\sum Y_{.j} + x_i)]^2 \le \sigma_2^2 \quad (6)$$

$$\frac{1}{rt}\left\{\sum(Y_{ij})^2 + x_i^2\right\} - [\frac{1}{r}(\sum Y_{ij} + x_i)]^2 \le \sigma_T^2 \quad (7)$$

$$x_i \ge 0$$

Where σ_1^2, σ_2^2 and σ_T^2 are the first source variability, second source variability and the total variability.

2. Flaws in Bhad and Ahmed (2012) and their corrections

In model M1 there is no distinction between the subscript used for the missing values and sources of variations. Also, the i[th] source of variation is not restricted to where the missing value is situated as illustrated in the example. If an observation is missing in the l[th] level of variation source m, then constraints (2) could have been expressed as follows:

$$Var(S_{ml}) \le \sigma_{ml}^2, \; m = 1, 2, \ldots, n \quad (8)$$

Where S_{ml} is l[th] level of source m where a missing value occurs, and σ_{ml}^2 is the variance of the l[th] level of source m excluding the missing value.

With the above definitions, the model M1 should have been written as

M3: $$\text{Minimize } \{f(x_i) = SSE\}$$

Subject to:

$$\text{Variance}(S_{ml}) \le \sigma_{ml}^2, \; m = 1, 2, \ldots, n; \text{ for each } l.$$

$$x_i \ge 0, \; i = 1, 2, \ldots, k$$

The restricted model M2 should have been formulated as follows:

Let Y_{ij} be the value in the i[th] level of source 1 and jth level of source 2, where i = 1, 2, . . . , t; j = 1, 2, . . . , r. Suppose the value Y_{pq} is missing, where p is the level of source 1 and q is the level of source 2, then the variance of the p[th] level of source 1, qth level of source 2 and the total are given, respectively, as

$$\mathrm{Var}(S_{1p}) = \frac{1}{r}\left\{\sum_{j\neq q}^{r} Y_{pj}^2 + Y_{pq}^2\right\} - \left[\frac{1}{r}\left\{\sum_{j\neq q}^{r} Y_{pj} + Y_{pq}\right\}\right]^2 \quad (9)$$

$$\mathrm{Var}(S_{2q}) = \frac{1}{t}\left\{\sum_{i\neq p}^{t} Y_{iq}^2 + Y_{pq}^2\right\} - \left[\frac{1}{t}\left\{\sum_{i\neq p}^{t} Y_{iq} + Y_{pq}\right\}\right]^2 \quad (10)$$

$$\mathrm{Var}(\mathrm{Total}) = \frac{1}{r}\left\{\sum_{i\neq p}^{t}\sum_{j\neq q}^{r} Y_{ij}^2 + Y_{pq}^2\right\} - \left[\frac{1}{r}\left\{\sum_{i\neq p}^{t}\sum_{j\neq q}^{r} Y_{ij} + Y_{pq}\right\}\right]^2 \quad (11)$$

The sum of squares of errors (SSE) should been expressed as follows:

$$SSE = Y_{pq}^2\left\{1 - \frac{1}{t} - \frac{1}{r} + \frac{1}{rt}\right\} - 2Y_{pq}\left\{\frac{Y'_{.q}}{t} + \frac{Y'_{p.}}{r} - \frac{Y'_{..}}{rt}\right\} + C \quad (12)$$

Where

$Y'_{p.}$ = Level p total of source 1 variation excluding the missing value Y_{pq}

$Y'_{.q}$ = Level q total of source 2 variation excluding the missing value Y_{pq}

$Y'_{..}$ = Grand total excluding the missing value Y_{pq}

C = Terms independent of the missing value

The corrected version of mathematical programming model (M2) is therefore:

M4: $$\mathrm{Min}\{SSE = Y_{pq}^2\left\{1 - \frac{1}{t} - \frac{1}{r} + \frac{1}{rt}\right\} - 2Y_{pq}\left\{\frac{Y'_{.q}}{t} + \frac{Y'_{p.}}{r} - \frac{Y'_{..}}{rt}\right\}\} \quad (13)$$

Subject to:

$$\frac{1}{r}\left\{\sum_{j\neq q}^{r} Y_{pj}^2 + Y_{pq}^2\right\} - \left[\frac{1}{r}\left\{\sum_{j\neq q}^{r} Y_{pj} + Y_{pq}\right\}\right]^2 \le \sigma_{1p}^2 \quad (14)$$

$$\frac{1}{t}\left\{\sum_{i\neq p}^{t} Y_{iq}^2 + Y_{pq}^2\right\} - \left[\frac{1}{t}\left\{\sum_{i\neq p}^{t} Y_{iq} + Y_{pq}\right\}\right]^2 \le \sigma_{2q}^2 \quad (15)$$

$$\frac{1}{r}\left\{\sum_{i\neq p}^{t}\sum_{j\neq q}^{r} Y_{ij}^2 + Y_{pq}^2\right\} - \left[\frac{1}{r}\left\{\sum_{i\neq p}^{t}\sum_{j\neq q}^{r} Y_{ij} + Y_{pq}\right\}\right]^2 \le \sigma_{T}^2 \quad (16)$$

$$Y_{pq} \geq 0 \tag{17}$$

3. Deriving Computational formula for estimating two missing values

Suppose two values Y_{rs} and Y_{pq} are missing in a randomized complete block design. We estimate the missing values by solving the unconstrained optimization problem

$$\text{Mininmize } \{SSE = SST_0 - SSB - SST = f(Y_{rs}, Y_{pq})\} \tag{18}$$

where, SSE, SST_0, SSB and SST are the sum of squares of errors, sum of squares of total, sum of squares of blocks, and sum of squares of treatments, respectively.

$$SSE = \left(\sum_{i=1}^{b}\sum_{j=1}^{k} Y_{ij}^2 - \frac{Y_{..}^2}{bk}\right) - \left(\frac{1}{k}\sum_{i=1}^{b} Y_{i.}^2 - \frac{Y_{..}^2}{bk}\right) - \left(\frac{1}{b}\sum_{j=1}^{k} Y_{.j}^2 - \frac{Y_{..}^2}{bk}\right)$$

$$= \sum_{i\neq r,p}^{b}\sum_{j\neq s,q}^{k} Y_{ij}^2 + Y_{rs}^2 + Y_{pq}^2 - \frac{1}{k}\sum_{i\neq r,p}^{b} Y_{i.}^2 - \frac{Y_{r.}^2}{k} - \frac{Y_{p.}^2}{k} - \frac{1}{b}\sum_{j\neq s,q}^{k} Y_{.j}^2 - \frac{Y_{.s}^2}{b} - \frac{Y_{.q}^2}{b} + \frac{Y_{..}^2}{bk}$$

$$= \sum_{i\neq r,p}^{b}\sum_{j\neq s,q}^{k} Y_{ij}^2 + Y_{rs}^2 + Y_{pq}^2 - \frac{1}{k}\sum_{i\neq r,p}^{b} Y_{i.}^2 - \frac{(Y_{r.}^{/} + Y_{rs})^2}{k} - \frac{(Y_{p.}^{/} + Y_{pq})^2}{k}$$

$$-\frac{1}{b}\sum_{j\neq s,q}^{k} Y_{.j}^2 - \frac{(Y_{.s}^{/} + Y_{rs})^2}{b} - \frac{(Y_{.q}^{/} + Y_{pq})^2}{b} + \frac{(Y_{..}^{/} + Y_{rs} + Y_{pq})^2}{bk}$$

$$= R + Y_{rs}^2 + Y_{pq}^2 - \frac{(Y_{r.}^{/} + Y_{rs})^2}{k} - \frac{(Y_{p.}^{/} + Y_{pq})^2}{k} - \frac{(Y_{.s}^{/} + Y_{rs})^2}{b} - \frac{(Y_{.q}^{/} + Y_{pq})^2}{b}$$

$$+ \frac{(Y_{..}^{/} + Y_{rs} + Y_{pq})^2}{bk} \tag{19}$$

Where $Y_{r.}$', $Y_{p.}$', $Y_{.s}$', $Y_{.q}$' and $Y_{..}$' are the respective totals excluding the missing values; R represent terms not involving missing values.

Differentiating SSE in equation (19) with respect to Y_{re} and Y_{pq} respectively and setting each equal to zero we obtain

$$(bk - b - k + 1)Y_{rs} + Y_{pq} = bY_{r.}^{/} + Y_{.s}^{/} - Y_{..}^{/} \tag{20}$$

$$Y_{rs} + (bk - b - k + 1)Y_{pq} = bY_{p.}^{/} + Y_{.q}^{/} - Y_{..}^{/} \tag{21}$$

Now solving equations (20) and (21) simultaneously we obtain

$$Y_{rs} = \frac{(b-1)(k-1)(bY_{r.}^{/} + kY_{.s}^{/} - Y_{..}^{/}) - (bY_{p.}^{/} + kY_{.q}^{/} - Y_{..}^{/})}{[(b-1)(k-1)]^2 - 1} \tag{22}$$

$$Y_{pq} = \frac{(b-1)(k-1)(bY'_{p.} + kY'_{.q} - Y'_{..}) - (bY'_{r.} + kY'_{.s} - Y'_{..})}{[(b-1)(k-1)]^2 - 1} \tag{23}$$

Example

We apply our formular to the example in Bhat and Ahmed (2012).

The table below shows a randomized complete block design with two missing values Y_{22} and Y_{35}.

Block	**Treatment**						**$Y_{i.}$**
	1	**2**	**3**	**4**	**5**	**6**	
1	18.5	15.7	16.2	14.1	13.0	13.6	91.1
2	11.7	Y_{22}	12.9	14.4	16.9	12.5	68.4 + Y_{22}
3	15.4	16.6	15.5	20.3	Y_{35}	21.5	89.3 + Y_{35}
4	16.5	18.6	12.7	15.7	16.5	18.0	98.0
$Y_{.j}$	62.1	50.9 + Y_{22}	57.3	64.5	46.4 + Y_{35}	65.6	346.8 + Y_{22} + Y_{35}

In our formulars (equations 20 and 21), b = 4, k = 6, r = 2, s = 2, p = 3, q = 5, $Y_{2.}'$ = 68.4, $Y_{3.}'$ = 89.3, $Y_{.2}'$ = 50.9, $Y_{.5}'$ = 46.4, $Y_{..}'$ = 346.8.

So,

$$Y_{22} = \frac{(b-1)(k-1)(bY'_{2.} + kY'_{.2} - Y'_{..}) - (bY'_{3.} + kY'_{.5} - Y'_{..})}{[(b-1)(k-1)]^2 - 1}$$

$$= \frac{15(4 \text{ x } 68.4 + 6 \text{ x } 50.9 - 346.8) - (4 \text{ x } 89.3 + 6 \text{ x } 46.4 - 346.8)}{224} = 14.3$$

$$Y_{35} = \frac{(b-1)(k-1)(bY'_{3.} + kY'_{.5} - Y'_{..}) - (bY'_{2.} + kY'_{.2} - Y'_{..})}{[(b-1)(k-1)]^2 - 1}$$

$$= \frac{15(4 \text{ x } 89.3 + 6 \text{ x } 46.4 - 346.8) - (4 \text{ x } 68.4 + 6 \text{ x } 50.9 - 346.8)}{224} = 18.3$$

5. Comparison of two methods

The results obtained by the two methods are summarized in the table below:

Missing Value	Bhad and Ahmed (2012)	Formular
Y_{22}	14.3	14.3
Y_{35}	17.0	18.3
SSE	80.7	79.6

5. Conclusions

As can be seen in the above table the estimate of missing values obtained by our formular produces a lower sum of squares of errors and hence a better estimate than that of Bhad and Ahmed (2012).