

Measure of overall regression sum of squares of symmetric randomized complete block design with a lost observation

Kittiwat Sirikasemsuk^{1*}

¹Department of Industrial Engineering, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
*Corresponding author E-mail: kittiwat.sirikasemsuk@gmail.com

Abstract

A randomized complete block design (RCBD) is useful for analyzing a treatment variable and one block variable under the condition where experimental units are limited. The RCBD is assumed that there is no interaction between the treatment variable and the block variable. This paper considered the symmetric randomized complete block design (SRCBD) with t treatments and t blocks, when a lost value occurs in the experiments. For the analysis of variance for the unbalanced data, the ready-made formulae were not provided in the past. The SRCBD with a lost value was analyzed by means of the fundamental underlying linear regression model in order to determine the reliable mathematical formulae for the fitted parameters and the overall regression sum of squares of experimental data. It is noted that all possible parameters are considered in the overall regression sum of squares which will be helpful for the analysis of variance through the exact approach (the model comparison approach) at a later stage.

Keywords: Design of Experiment; Randomized Complete Block Design; Lost Observation; Overall Regression Sum of Squares

1. Introduction

Randomization, replication, and blocking are the three primary concepts of the planning for the classical DOE based on Fisher's principles [1]. Experimentation plays an important role in the design and improvement of product and production, mainly in order to determine the most influential variables on the response variable and set the influential variables so that the value of response variable is near the desired nominal value. The randomized complete block design (RCBD), the balanced incomplete block design (BIBD), and the Latin square design (LSD) are three well-known examples of the use of the blocking principle. Both RCBD and BIBD support the examination of experiments for one suspected factor (treatment variable) and one blocked nuisance (known and controllable variable). The BIBD is based on the background information on the RCBD where all treatment-unit combinations of the RCBD can be run in each block. However, the BIBD allows experimenters to ignore some experimental observations in each block and each treatment of suspected factor in the planning stage owing to the limited number of facilities, apparatus or things used in experiments (i.e., material, animals, plants, food and so on), including the limitation of the high experimental cost [2]. The BIBD can be applied to the case that the large number of treatments is preferable with the incomplete observations in each treatment. In addition, the symmetric randomized complete block design (SRCBD) is symmetric if a number of treatments are equal to a number of blocks [3].

Besides using the classical DOE, a number of experimenters turn to Taguchi method proposed by Dr. Genichi Taguchi in the 1980s. This method is the analysis of a robust design via the signal-to-noise ratio. Taguchi method is designed to try to examine the effect of the large number of potential variables on the response variable. For more details on these examples of the implementation, readers are referred to the papers of Manivannan et al. [4] and Ngo et al. [5].

In a real scientific test under certain conditions, the experimenters might face a difficult situation in which a set of experimental observations is not complete. The incomplete-observation situation can be commonly divided into two situations: (1) the initial intention to cause the incomplete observations due to a limitation on the number of experimental units, (2) the accidental situation.

The first situation can be either the existence of balanced characteristic or unbalanced arrangement. For instance, Youden [6], Yates [7], and Ai et al. [8], respectively proposed the Youden square design, the BIBD, and the balanced incomplete Latin square design (BILSD). Such a balanced arrangement can help make the analysis of the variances (ANOVA) easier with the simple formulae to determine the treatment and error sums of squares. In the second situation which might occur from bad control of some variables, the reading values from experiment are abnormal or not observed. Hence, their values might be cut from a set of observations, leading to the unbalanced or asymmetrical arrangement. It is important to note that there is no certain formula for the analysis of the variances (ANOVA) in the incomplete-observation experimental design.

A rich literature review on the missing plot technique with non-iterative and iterative methods according to the original paper of Allan and Wishart [9] and their extension is given by Little and Rubin [10]. In their missing plot technique, the differential calculus is used to determine the lost observations with the minimal error sum of squares. The estimates of the lost observations for the vacant cells usually cause the treatment sum of squares with an upward bias. Hence, after using the missing-plot technique, the next process is the determination of the bias. Afterward, this bias is subtracted from the initial treatment sum of squares. In addition, Baird and Kramer [11] considered the BIBD consisting of the additional lost observations. They focused on the estimation of the additional lost observations by minimizing the error sum of squares.

The exact approach (the model comparison approach) relying on the general linear model is the method in which the determination of the lost observations is not made. This approach is what Montgomery [1] calls ‘the general regression significance test’. Next, the analysis of covariance technique was applied to address the issue of incomplete experimental designs associated with the missing-plot technique, e.g. Coons [12], Cochran [13], and Wilkinson [14]. The earliest paper with a reference to the analysis of covariance technique in the experiment belongs to Bartlett [15]. Recently, Ogbonnaya and Uzochukwu [16] provided the explicit formulae for the estimates of the single or several lost data for one-factor analysis of covariance with one covariate. The use of the analysis of covariance technique helps give the unbiased estimates of the error and treatment sums of squares.

Above all, methodologies, except the exact approach, must estimate the lost observations for the vacant cells. With a common sense view, the lost observations should not be estimated because their values are not derived from the actual and appropriate experiments. Hence, this research encouraged the use of the exact approach with the general linear model in order to tackle the problems of incomplete experimental design. Recently, the expression of the overall regression sum of squares for the LSD with the one lost data was developed by Sirikasemsuk [17] and Sirikasemsuk and Leerojanaprapa [18]; while the case of the two lost data was given by Sirikasemsuk and Leerojanaprapa [19].

The current research dealt with the SRCBD consisting of one lost observation. Our aim was to provide the explicit and mathematical formulae for the estimates of the fitted parameters and the overall regression sum of squares. It is also noted that the overall regression sum of squares for the full effect model helps calculate the treatment sum of squares for the analysis of the variance. In this research, the SRCBD with any t treatments (or t blocks) is considered. The examples of SRCBD can be shown in Fig. 1.

The notations and their respective definitions used in this research are as follows:

- y_{ij} the ij^{th} observation taken under row (treatment) i , and column j
- i index of rows or treatments ($i = 1, 2, 3, \dots, t$)
- j index of columns ($j = 1, 2, 3, \dots, t$)
- t the number of treatments (or blocks)
- μ the parameter of the overall mean
- ω_i the row effects of level i of blocking variable
- λ_j the treatment effects of level j of treatment variable
- $\hat{\mu}$ the estimate of the parameter of μ
- $\hat{\omega}_i$ the estimate of the parameter of the i^{th} row effect or treatment effect
- $\hat{\lambda}_j$ the estimate of the parameter of the j^{th} column effect
- $y_{..}$ the grand total

- $y_{i.}$ the i^{th} row total or treatment total
- $y_{.j}$ the j^{th} column total
- r index of the row in which the observation is missing
- c index of the column in which the observation is missing
- $R(\mu, \omega, \lambda)$ the regression sum of squares for the full effect model of y_{ij}

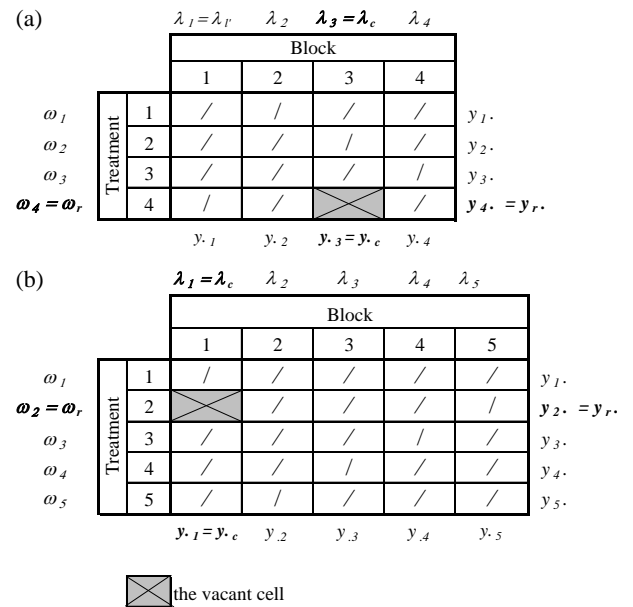


Fig.1: Two sample configurations of the SRCBD with one lost observation: (a) $t = 4$ (b) $t = 5$

The organization of the rest of this research is as follows: Section 2 provides the set of least square normal equations and the estimates of the fitted parameters. In Section 3, the ready-made formula of the overall regression sum of squares is derived to make simplicity. The concluding remarks are provided in Section 4.

2. Estimates of full-effect-model parameters

Based on the an additive effects model, which does not include the block by the factor interaction terms, the formal linear-model structure of the full effect model for y_{ij} of the RCBD (or the SRCBD) is

$$y_{ij} = \mu + \omega_i + \lambda_j + \varepsilon_{ij} \begin{cases} i = 1, 2, 3, \dots, t \\ j = 1, 2, 3, \dots, t \end{cases} \quad (1)$$

where ε_{ij} is random or unexplained error in the value of the response variable within each treatment–block combination and is assumed to be normally distributed with a mean of zero and a certain variance.

According to the work of Rencher and Schaalje [20], the overall regression sum of squares for the full effect model of y_{ij} is adapted as

$$R(\mu, \omega, \lambda) = \hat{\mu}y_{..} + \sum_{i=1}^t \hat{\omega}_i y_{i.} + \sum_{j=1}^t \hat{\lambda}_j y_{.j} \quad (2)$$

To determine the estimated values of parameters in the full effect model, the least square normal equations for the $t \times t$ SRCBD design with one lost value are:

$$\mu : (t^2 - 1)\hat{\mu} + t \sum_{i=1, i \neq r}^t \hat{\omega}_i + (t-1)\hat{\omega}_r + t \sum_{j=1, j \neq c}^t \hat{\lambda}_j + (t-1)\hat{\lambda}_c = y_{..} \quad (3)$$

$$\omega_r : (t-1)\hat{\mu} + (t-1)\hat{\omega}_r + \sum_{j=1, j \neq c}^t \hat{\lambda}_j = y_r. \quad (4)$$

$$\lambda_c : (t-1)\hat{\mu} + \sum_{i=1, i \neq r}^t \hat{\omega}_i + (t-1)\hat{\lambda}_c = y_c \quad (5)$$

$$\omega_i : t\hat{\mu} + t\hat{\omega}_i + \sum_{j=1}^t \hat{\lambda}_j = y_i, \text{ when } i \neq r \quad (6)$$

$$\lambda_j : t\hat{\mu} + \sum_{i=1}^t \hat{\omega}_i + t\hat{\lambda}_j = y_j, \text{ when } j \neq c \quad (7)$$

In the RCBD, there are the following restricted assumptions:

$$\sum_{All\ i}^K \hat{\omega}_i = 0 \quad (8)$$

and

$$\sum_{All\ j}^t \hat{\lambda}_j = 0 \quad (9)$$

With Eqs. (8) and (9), the normal equations (3)-(7) can be condensed into the following equations:

$$\mu : (t^2 - 1)\hat{\mu} - \hat{\omega}_r - \hat{\lambda}_c = y_{..} \quad (10)$$

$$\omega_r : (t-1)\hat{\mu} + (t-1)\hat{\omega}_r - \hat{\lambda}_c = y_r. \quad (11)$$

$$\lambda_c : (t-1)\hat{\mu} - \hat{\omega}_r + (t-1)\hat{\lambda}_c = y_c \quad (12)$$

$$\omega_i : t\hat{\mu} + t\hat{\omega}_i = y_i, \text{ when } i \neq r \quad (13)$$

$$\lambda_j : t\hat{\mu} + t\hat{\lambda}_j = y_j, \text{ when } j \neq c \quad (14)$$

The estimates of the fitted parameters of Eq. (1) can be solved and determined as the following Proposition 1.

Proposition 1: In the SRCBD with one lost value, the estimates of the fitted parameters in the full effect model of $y_{..}$ can be determined as

$$\hat{\mu} = \frac{(t-2)y_{..} + y_r + y_c}{t(t-1)^2} \quad (15)$$

$$\hat{\omega}_r = (t-1)\hat{\mu} + \frac{y_r - y_{..}}{t} \quad (16)$$

$$\hat{\lambda}_c = (t-1)\hat{\mu} + \frac{y_c - y_{..}}{t} \quad (17)$$

$$\hat{\omega}_i = \frac{y_i}{t} - \hat{\mu} \text{ when } i \neq r \quad (18)$$

and

$$\hat{\lambda}_j = \frac{y_j}{t} - \hat{\mu} \text{ when } j \neq c \quad (19)$$

Proof. Multiplying $(t-2)$ on both sides of Eq. (10), then adding Eqs. (11) and (12), and rearranging, the parameter estimate of μ

is expressed in Eq. (15). The fitted parameters $\hat{\omega}_r$ and $\hat{\lambda}_c$ in Eqs. (16) and (17) can be easily solved from Eqs. (11) and (12). When $i \neq r$ and $j \neq c$, it is noted that the fitted parameters $\hat{\omega}_i$ and $\hat{\lambda}_j$ in Eqs. (18) and (19) can be easily solved from Eqs. (13) and (14). This completes the proof. \square

3. Overall regression sum of squares

The overall regression sum of squares can be solved and determined as the following Proposition 2.

Proposition 2: In the SRCBD with one lost value, the overall regression sum of squares for the full effect model of y_{ij} can be determined as

$$R(\mu, \omega, \lambda) = \frac{\sum_{All\ i=1}^t y_i^2 + \sum_{All\ j=1}^t y_j^2}{t} + \frac{(y_r + y_c - y_{..})[t(y_r + y_c) + (t-2)y_{..}]}{t(t-1)^2} \quad (20)$$

Proof. Substituting Eqs. (16), (17), (18) and (19) into Eq. (2), we obtain

$$R(\mu, \omega, \lambda) = \frac{\sum_{All\ i=1}^t y_i^2 + \sum_{All\ j=1}^t y_j^2}{t} + t\hat{\mu}(y_r + y_c) - \frac{y_{..}}{t}(y_r + y_c) - \hat{\mu}y_{..} \quad (21)$$

Substituting Eq. (15) in Eq. (21) together with the algebraic simplification yields Eq. (20). This completes the proof. \square

An attracting illustration which was laid out in a 5×5 SRCBD as shown in Table 1, is adapted. There were five different investigators (treatments) on five separate days.

The overall regression sum of squares for the full model can be easily calculated by Eq. (21) and is equal to 11325.823. It is noted that the overall regression sum of squares will lead to the calculation for the unbiased ANOVA table later.

Table 1: Example

Investigator (Treatment Variable)	Day (Block Variable)					Total
	1	2	3	4	5	
1	22.1	18.6	23.0	24.3	17.1	$y_1 = 105.1$
2	23.5	16.5	18.7	22.0	---	$y_2 = 80.7$
3	17.4	23.8	22.8	23.9	20.0	$y_3 = 107.9$
4	20.3	23.4	25.9	18.7	24.2	$y_4 = 112.5$
5	25.7	24.8	18.9	20.6	24.6	$y_5 = 114.6$
	$y_1 = 109$	$y_2 = 107.1$	$y_3 = 109.3$	$y_4 = 109.5$	$y_5 = 85.9$	$y_{..} = 520.8$

4. Conclusion

The literature showed that there was no ready-made formula to solve the RCBD with a lost value through the exact approach with a general regression significance test in which no estimate of the lost values was made. In the unbalanced designs, the usual formulae for determination of all sums of squares cannot be used for the ANOVA table. The exact approach with the general linear model was preferable to tackle the problems of incomplete block design in the paper. The SRCBD of any number of treatments consisting of one unobserved data with one replication was considered. This current paper used the least square normal equations to find the estimated full-model parameters and to provide the explicit and mathematical formula for the overall regression sum of squares for the full effect model, thereby simplifying the calculation process. The proposed formula was applicable to the one-lost-value SRCBD of any configuration. In addition, it is important to note that the overall regression sum of squares in this current paper will help construct the ANOVA table applicable to the SRCBD with one lost observation. In the future research, the overall regression sum of squares of the reduced effect model that ignores the treatment effects should be determined. The treatment sum of squares can also be derived from the difference between the regression sums of squares of the full effect model and the reduced effect model.

References

- [1] Montgomery DC (2008), Design and Analysis of Experiments, 5th edn. *John Wiley & Sons*.
- [2] Mead R, Gilmour SG, & Mead A (2012), Statistical Principles for the Design of Experiments: Applications to Real Experiments, Vol. 36, *Cambridge University Press*.
- [3] Kelechi AC (2012), Symmetric and unsymmetric balanced incomplete block designs: a comparative analysis. *International Journal of Statistics and Applications* 2, pp. 33-39.
- [4] Manivannan S, Arumugam R, Sudharsan NM, & Prasanna DS (2010), Taguchi based linear regression modeling of flat plate heat sink. *Journal of Engineering and Applied Sciences* 5, pp. 36-44.
- [5] Ngo MT, Hoang V, & Hoang VS (2017), Optimization cutting fluids and parameters for minimizing cutting force in fly-hobbing. *Journal of Engineering and Applied Sciences* 12, pp. 5212-5218.
- [6] Youden WJ (1937), Use of incomplete block replications in estimating tobacco-mosaic virus. *Contributions from Boyce Thompson Institute* 9, pp. 41-48.
- [7] Yates F (1936), Incomplete randomized blocks. *Annals of Eugenics* 7, pp. 121-140.
- [8] Ai M, Li K, Liu S, & Lin DK (2013), Balanced incomplete latin square designs. *Journal of Statistical Planning and Inference* 143, pp. 1575-1582.
- [9] Allan FE & Wishart J (1930), A method of estimating the yield of a missing plot in field experimental work. *Journal of Agricultural Science* 20, pp. 399-406.
- [10] Little RJA & Rubin DB (2002), Statistical Analysis with Missing Data, 2nd edn. *John Wiley & Sons*, New York.
- [11] Baird HR & Kramer CY (1960), Analysis of variance of a balanced incomplete block design with missing observations. *Journal of the Royal Statistical Society: Series C* 9, pp. 189-198.
- [12] Coons I (1957), The analysis of covariance as a missing plot technique. *Biometrics* 13, pp. 387-405.
- [13] Cochran WG (1957), Analysis of covariance: its nature and uses. *Biometrics* 13, pp. 261-281.
- [14] Wilkinson GN (1958), Estimation of missing values for the analysis of incomplete data. *Biometrics* 14, pp. 257-286.
- [15] Bartlett MS (1937), Some examples of statistical methods of research in agriculture and applied biology. *Journal of the Royal Statistical Society* 4, pp. 137-183.
- [16] Ogbonnaya CE & Uzochukwu EC (2016), Estimation of missing data in analysis of covariance: a least-squares approach. *Communications in Statistics - Theory and Methods* 45, pp. 1902-1909.
- [17] Sirikasemsuk K (2016), One missing value problem in latin square design of any order: regression sum of squares. Proceedings of the Joint 8th International Conference on Soft Computing and Intelligent Systems and 17th International Symposium on Advanced Intelligent Systems (SCIS&ISIS), pp. 142-147.
- [18] Sirikasemsuk K & Leerojanaprapa K (2017), One missing value problem in latin square design of any order: exact analysis of variance. *Cogent Engineering* 4, pp. 1411222.
- [19] Sirikasemsuk K & Leerojanaprapa K (2017), Analysis of two-missing-observation 4x4 latin squares using the exact approach. *Advances in Intelligent Systems and Computing* 566, P. Meesad P, Sunantha S, Unger H, Eds. Springer, pp. 69-81.
- [20] Rencher AC & Schaalje GB (2008), Linear Models in Statistics, 2nd edn. *John Wiley & Sons*.