One Missing Value Problem in Latin Square Design of Any Order: Exact Analysis of Variance

Kittiwat SirikasemsukKanogkan Leerojanaprapa

Accepted Manuscript Version

Publisher: Cogent OA

Journal: *Cogent Engineering*

DOI: http://doi.org/10.1080/23311916.2017.1411222

One Missing Value Problem in Latin Square Design of Any Order: Exact Analysis of Variance

Kittiwat Sirikasemsuk, Kanogkan Leerojanaprapa

Kittiwat Sirikasemsuk[1] and Kanogkan Leerojanaprapa[2]

[1] Department of Industrial Engineering, Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
kittiwat.sirikasemsuk@gmail.com
[2] Statistics Department, Faculty of Science,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
kanogkan.le@kmitl.ac.th

Abstract. This research proposes a simplified exact approach based on the general linear model for solving the $K \times K$ Latin square design with one replicate and one missing value, given the lack of ready-made mathematical formulas for the sub-variance. Under the proposed scheme, the effects of the potential variable were determined by means of the regression sums of squares under the full and reduced treatment models. The mathematical expressions could be applied to the Latin square design with one missing value of any order. Moreover, the treatment, row and column sums of squares are unbiased.

Keywords: Experimental Design, Incomplete Latin Square, ANOVA, General Linear Model

1 Introduction

In science and engineering, design of experiments (DOE) refers to the experimental situations or strategies for analysis of quantitative responses associated with the experimental units. DOE is classified into various types, including the classical DOE based on Fisher's principles, Shainin experiment, Taguchi experiment. Specifically, the DOE based on Fisher's principles involves randomization, replication and blocking [1]. In the design and improvement of products and production, the role of experimentation is to identify the influencing factors (determinants) of the response variable and manipulate the determinants such that the response variable outcome closely resembles the desired nominal value. In fact, Fisher's classical DOE is a form of statistical hypothesis testing under the analysis of variance (ANOVA) [2]. Meanwhile, ANOVA is defined as a collection of statistical procedures to compare the between-group variation with the within-group variation [1].

A Latin square design (LSD) is an efficient design of experiments for three factors, whereby only one factor is of primary interest (i.e. the potential variable) while the other two (the nuisance variables or factors) are blocked to restrain extraneous variability in experimental units. The word 'Latin square design' is abbreviated to 'LSD' in this research. Latin letters are used to symbolize the level of the factor of primary interest. In the LSD, the levels of the two nuisance variables are identified with the rows and columns of a two-way table; every level of the factor of primary interest appears once in each column and once in each row; and the two-factor and three-factor interaction effects are assumed non-existent. Besides the randomized complete block design (RCBD), in which the effect of a single nuisance variable is blocked, the LSD also utilizes the blocking technique to separate the variations of nuisance variables from the experimental error. Unlike in the LSD, in the Latin rectangle design [3] the numbers of columns and rows (blocks) are not identical for the two nuisance factors, and the Latin letters in each row (or column) can be replicated. In [4], the Youden square design or the distinct Latin rectangle design was proposed whereby the number of blocks on one side is greater than the other side's, and the number of treatments (Latin letters) is equal to the number of blocks of the former.

In a real scientific test under certain conditions, 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 occur the incomplete observations due to a limitation on the number of experimental units, i.e., material units, articles, or subjects, (2) the accidental situation. The first situation can be the existence of balanced characteristic or unbalanced arrangement. For instance, Youden [4], Yates [5], and Ai et al. [6], respectively proposed the Youden square design, the balanced incomplete block design (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 The work of Allan and Wishart [7] seems to be the earliest paper specifically considering the analysis of incomplete-data problem by means of the differentiation based on the overall mean. In [8], [9], the non-iterative and iterative missing plot techniques were proposed whereby the differential calculus was utilized to determine the missing experimental data with minimal error sum of squares. The estimates of the missing experimental data however contribute to an upward bias of the treatment sum of squares. Thus, the bias is determined and subtracted from the initial treatment sum of squares [10]. In [11], [12], [13], the analysis of covariance (ANCOVA) technique was proposed for solving the incomplete-data experimental designs. In fact, the earliest paper with a reference to the ANCOVA was Bartlett's [14].

Table 1 tabulates existing methods to solve the incomplete-data experimental problems. However, the single imputation methods based on the mean (or mode) substitution, listwise deletion and pairwise deletion are excluded.

Table 1. Existing methods to solve the incomplete-data experimental problems

Method	Description	Author
Missing plot technique by minimizing the error sum of squares with non-iterative method	- Differentiating the estimated parameter of the overall mean with respect to each missing value	- Allan and Wishart [7]
	- Differentiating the error sum of squares to each missing value (when only one observation is missing)	- Yates [8]
	- General method for estimating several missing values in Latin square design	- Kramer and Glass [15]
	- Non-iterative Rubin method	- Rubin [16]
Missing plot technique with iterative method	- Iterative Yates method (based on the work of Allan and Wishart [7]) when more than one observations are missing)	- Yates [8]
	- Healy-Westmacott method based on regression imputation	- Healy and Westmacott [17]
Exact approach with general linear model	General regression significance test	Montgomery [1]
Analysis of covariance (ANCOVA) technique	A combination of regression analysis and ANOVA consisting of the covariate	Coons [11], Cochran [12], and Wilkinson [13]
Expectation maximization algorithm (EM Algorithm)	Iterative method with maximum likelihood estimation	Dempster et al. [18]
Multiple imputation (MI) method	A combination of raw maximum likelihood and EM method	Rubin [19]

Many recent research studies considered aspects of combinatorics, examples of which were the studies on the construction of the orthogonal Latin squares by Zhang [20] and Donovan and Yazıcı [21]; and the studies on the completability of the incomplete Latin squares from the partial Latin squares by Euler [22] and Casselgren and Häggkvist [23].

In Table 1, all the methods, except the exact approach, must estimate the missing observations. As a matter of fact, the missing observations should never be estimated because the estimate values are not experiment-based. Thus, it is advisable that the exact approach with the general linear model be adopted to solve the incomplete-data experimental design problems [1, 9]. Specifically, this research proposes a simplified exact approach (the general regression significant test) for the $K \times K$ LSD with one replicate and one missing experimental data, where K is the order of the LSD.

The organization of this research is as follows: Section 1 is the introduction. Section 2 details the general ANOVA table for a complete LSD with $K \times K$ order and the components. Section 3 deals with a $K \times K$ LSD with one missing data, the estimated parameter values of the full effect model and the regression sum of squares, while Section 4 concerns those of the reduced-treatment effect model and the regression sum of squares. Section 5 derives the simplified formulas of the sums of squares. The concluding remarks are provided in Section 6. The notations are provided in the appendix.

2 Analysis of Variance in Complete *K* × *K* Latin Square Design

The full effect model of y_{ijk}, given the complete $K \times K$ LSD, is expressed as

$$y_{ijk} = \mu + \omega_i + \tau_j + \lambda_k + \varepsilon_{ijk} \qquad (1)$$

where ε_{ijk} is independently, identically and normally distributed, i.e. $\varepsilon_{ijk} \sim \mathrm{N}(0, \sigma^2)$. Table 2 presents an example of the LSD with $K \times K$ order whose components are summarized and tabulated using an ANOVA table, as shown in Table 3.

The sums of squares for a Latin square experiment are expressed as

$$SS_{tr} = \sum_{j=1}^{K} (\hat{\tau}_j)^2 = \sum_{j=1}^{K} \left(\bar{y}_{\cdot j \cdot} - \bar{y}_{\cdots}\right)^2 \qquad (2)$$

$$SS_{row} = \sum_{i=1}^{K} (\hat{\omega}_i)^2 = \sum_{i=1}^{K} \left(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}\right)^2 \qquad (3)$$

$$SS_{column} = \sum_{k=1}^{K} (\lambda_k)^2 = \sum_{k=1}^{K} \left(\bar{y}_{\cdot\cdot k} - \bar{y}_{\cdots}\right)^2 \qquad (4)$$

$$SS_{total} = \sum_{i=1}^{K} \sum_{k=1}^{K} \left(y_{i(j)k} - \bar{y}_{\cdots}\right)^2 \qquad (5)$$

$$SS_E = SS_{total} - (SS_{tr} + SS_{row} + SS_{column}) \qquad (6)$$

Table 2. The complete Latin square design of $K \times K$ order

Blocking	Blocking Variable 2						
Variable 1	**1**	**2**	**3**	**.**	**.**	**.**	***K***

1	A = y_{111}	B = y_{122}	C = y_{133}	.	.	.	Z* = y_{1KK}
2	B = y_{221}	C = y_{232}	D = y_{243}	.	.	.	A = y_{21K}
3	C = y_{331}	D = y_{342}	E = y_{353}	.	.	.	B = y_{32K}
.			.				
.			.				
.			.				
K	Z* = y_{KK1}	A = y_{K12}	B = y_{K23}	.	.	.	Y* = $y_{K(K-1)K}$

*Assume that the letters Z and Y represent the last and second-to-last levels of treatment.

Table 3. The ANOVA table for the complete $K \times K$ LSD

Source of Variation	Degrees of Freedom (*df*)	Sum of Squares (*SS*)	Mean Square (*MS*)	F_0
Treatments	$K-1$	SS_{tr}	$MS_{tr} = \frac{SS_{tr}}{K-1}$	$F_{tr} = MS_{tr}/MS_E$
Rows (Blocking Variable 1)	$K-1$	SS_{row}	$MS_{row} = \frac{SS_{row}}{K-1}$	$F_{row} = MS_{row}/MS_E$
Columns (Blocking Variable 2)	$K-1$	SS_{column}	$MS_{column} = \frac{SS_{column}}{K-1}$	$F_{column} = MS_{column}/MS_E$
Error	$(K-2)(K-1)$	SS_E	$MS_E = \frac{SS_E}{(K-2)(K-1)}$	
Total	K^2-1	SS_{total}		

3 Incomplete Latin Square Design and Regression Sum of Squares under the Full Model

For the missing-data Latin square designs, the sums of squares in Eqs. (2)-(4) are invalid. The general regression significance test could instead be applied to the incomplete Latin square design for analysis of variance. According to Montgomery [1], the computational formulas for the sums of squares of treatments, rows, columns and errors could respectively be expressed as

$$SS_{tr} = R(\mu,\omega,\tau,\lambda) - R(\mu,\omega,\lambda) \qquad (7)$$

$$SS_{row} = R(\mu,\omega,\tau,\lambda) - R(\mu,\tau,\lambda) \qquad (8)$$

$$SS_{column} = R(\mu,\omega,\tau,\lambda) - R(\mu,\omega,\tau) \quad (9)$$

$$SS_E = \sum_{i=1}^{K}\sum_{k=1}^{K} y_{i(j)k}^2 - R(\mu,\omega,\tau,\lambda) \quad (10)$$

where $R(\mu,\tau,\lambda)$ and $R(\mu,\omega,\tau)$ are the regression sums of squares of the reduced effect model of y_{ijk}, in which the effects of rows and columns are overlooked, respectively; and if one observation is missing, the degrees of freedom of SS_{total} and SS_E in Table 3 would respectively be $K^2 - 2$ and $K^2 - 3K + 1$.

Meanwhile, the theoretical regression sum of squares of the full effect model of y_{ijk} is expressed as

$$R(\mu,\omega,\tau,\lambda) = \hat{\mu}y_{...} + \sum_{i=1}^{K}\hat{\omega}_i y_{i..} + \sum_{j=1}^{K}\hat{\tau}_j y_{.j.} + \sum_{k=1}^{K}\hat{\lambda}_k y_{..k} \quad (11)$$

In [24], the estimated values of all parameters (Eq.(11)) of the incomplete LSD with one missing observation were derived and the regression sum of squares of the full effect model of y_{ijk} could be expressed as

$$R(\mu,\omega,\tau,\lambda) = \frac{\sum_{All\ i}^{K} y_{i..}^2 + \sum_{All\ j}^{K} y_{.j.}^2 + \sum_{All\ k}^{K} y_{..k}^2}{K} + \frac{(1-K)\left(2y_{...}^2 - y_{sum\_m}^2\right) + \left(y_{sum\_m} - 2y_{...}\right)^2}{K(K-1)(K-2)}$$

(12)

where $y_{sum\_m} = y_{r..} + y_{.m.} + y_{..c}$.

To find the treatment sum of squares (see Eq. (7)), it is assumed that the treatment effects (τ_j) are not considered in Eq. (1), i.e. $\tau_j = 0$ for all values of *j*. The estimated μ, ω_i, and λ_k will be substituted with $\hat{\mu}^{NT}$, $\hat{\omega}_i^{NT}$, and $\hat{\lambda}_k^{NT}$ instead of $\hat{\mu}$, $\hat{\omega}_i$ and $\hat{\lambda}_k$. With the treatment effects of a single factor is of primary interest ignored, this linear statistical model of y_{ijk} is referred to as "*the reduced-treatment effect model*" in this research. Thus, its regression sum of squares, $R(\mu,\omega,\lambda)$, can be expressed as Eq. (13).

$$R(\mu,\omega,\lambda) = \hat{\mu}^{NT} y_{...} + \sum_{i=1}^{K}\hat{\omega}_i^{NT} y_{i..} + \sum_{k=1}^{K}\hat{\lambda}_k^{NT} y_{..k} \quad (13)$$

The estimated model parameters, i.e. $\hat{\mu}$, $\hat{\omega}_i$, $\hat{\tau}_j$, $\hat{\lambda}_k$, $\hat{\mu}^{NT}$, $\hat{\omega}_i^{NT}$ and $\hat{\lambda}_k^{NT}$, will be later detailed in Section 4. It should be noted that the determination of the parameter estimates in $R(\mu,\tau,\lambda)$ and $R(\mu,\omega,\tau)$ is similarly carried out for

$R(\mu,\omega,\lambda)$ in Section 4. The expressions of $R(\mu,\tau,\lambda)$ and $R(\mu,\omega,\tau)$, including their parameter estimates, are not demonstrated in this research.

4 Estimated Values of All Parameters and Regression Sum of Squares under the Reduced-treatment Model

With the exact approach, it is necessary to find the estimates of the fitted values of the reduced-treatment effect model prior to $R(\mu,\omega,\lambda)$ according to Eq.(13). In addition, the parameter estimates for the reduced-treatment effect model can be divided into two categories: The first category refers to the parameter estimates directly influenced by the missing value, i.e. $\hat{\mu}^{NT}$, $\hat{\omega}_r^{NT}$ and $\hat{\lambda}_c^{NT}$ (see Proposition 1), while the second category consists of the remaining parameter estimates directly unaffected by the missing value, which can be derived and shown in Eqs. (22)-(23).

Proposition 1: In the $K \times K$ LSD with one missing experimental data, the estimates of the fitted parameters: $\hat{\mu}^{NT}$, $\hat{\omega}_r^{NT}$ and $\hat{\lambda}_c^{NT}$, in the reduced-treatment effect model can be determined by

$$\hat{\mu}^{NT} = \frac{(K-2)y_{...} + y_{r..} + y_{..c}}{K(K-1)^2} \qquad (14)$$

$$\hat{\omega}_r^{NT} = (K-1)\hat{\mu}^{NT} + \frac{y_{r..} - y_{...}}{K} \qquad (15)$$

and

$$\hat{\lambda}_c^{NT} = (K-1)\hat{\mu}^{NT} + \frac{y_{..c} - y_{...}}{K} \qquad (16)$$

Proof. Based on the restricted assumptions, i.e., $\sum_{All\ i}^{K} \hat{\omega}_i = 0$ and $\sum_{All\ k}^{K} \hat{\lambda}_k = 0$, the least square normal equations for the reduced-treatment effect model in which the parameter estimates are directly influenced by the missing-experimental-data position can be expressed as

$$\mu : \left(K^2-1\right)\hat{\mu}^{NT} - \hat{\omega}_r^{NT} - \hat{\lambda}_c^{NT} = y_{...} \qquad (17)$$

$$\omega_r : (K-1)\hat{\mu}^{NT} + (K-1)\hat{\omega}_r^{NT} - \hat{\lambda}_c^{NT} = y_{r..} \qquad (18)$$

$$\lambda_c : (K-1)\hat{\mu}^{NT} - \hat{\omega}_r^{NT} + (K-1)\hat{\lambda}_c^{NT} = y_{..c} \qquad (19)$$

Multiplying $(K-2)$ on both sides of Eq. (17), then adding Eqs. (18) and (19), and rearranging, the parameter estimate of μ is expressed in Eq. (14). The fitted parameters $\hat{\omega}_r^{NT}$ and $\hat{\lambda}_c^{NT}$ in Eqs. (15) and (16) can be easily solved from Eqs. (18) and (19). This completes the proof. ▯

In the second category of the reduced-treatment effect model in which the treatment effect is ignored, the normal equations can be expressed as

$$\omega_i : K\hat{\mu}^{NT} + K\hat{\omega}_i^{NT} + \sum_{k=1}^{K}\hat{\lambda}_k^{NT} = y_{i\cdot\cdot} \qquad (20)$$

$$\lambda_k : K\hat{\mu}^{NT} + \sum_{i=1}^{K}\hat{\omega}_i^{NT} + K\hat{\lambda}_k^{NT} = y_{\cdot\cdot k} \qquad (21)$$

where $i \neq r$ and $k \neq c$. The remaining parameter estimates are subsequently determined as

$$\hat{\omega}_i^{NT} = \frac{y_{i\cdot\cdot}}{K} - \hat{\mu}^{NT} \qquad (22)$$

$$\hat{\lambda}_k^{NT} = \frac{y_{\cdot\cdot k}}{K} - \hat{\mu}^{NT} \qquad (23)$$

where $i \neq r$, $k \neq c$; and $\hat{\mu}^{NT}$ in Eqs. (22) and (23) is substituted with Eq. (14). It is noted that the fitted parameters $\hat{\omega}_i^{NT}$ and $\hat{\lambda}_k^{NT}$ in Eqs. (22) and (23) can be easily solved from Eqs. (20) and (21).

Proposition 2: In the $K \times K$ LSD with one missing experimental data, the regression sum of squares for the reduced-treatment effect model of y_{ijk} can be expressed as

$$R(\mu,\omega,\lambda) = \frac{\sum_{All\ i=1}^{K} y_{i\cdot\cdot}^2 + \sum_{All\ k=1}^{K} y_{\cdot\cdot k}^2}{K} + \frac{(y_{r\cdot\cdot} + y_{\cdot\cdot c} - y_{\cdot\cdot\cdot})[K(y_{r\cdot\cdot} + y_{\cdot\cdot c}) + (K-2)y_{\cdot\cdot\cdot}]}{K(K-1)^2}$$

(24)

Proof. The determination of $R(\mu,\omega,\lambda)$ can be carried out in a similar fashion to that of $R(\mu,\omega,\tau,\lambda)$ in the paper of Sirikasemsuk [24] and is presented as below.

Substituting Eqs. (15), (16), (22) and (23) into Eq. (13), we obtain

$$R(\mu,\omega,\lambda) = \frac{\sum_{All\ i=1}^{K} y_{i\cdot\cdot}^2 + \sum_{All\ k=1}^{K} y_{\cdot\cdot k}^2}{K} + K\hat{\mu}^{NT}(y_{r\cdot\cdot} + y_{\cdot\cdot c}) - \frac{y_{\cdot\cdot\cdot}}{K}(y_{r\cdot\cdot} + y_{\cdot\cdot c}) - \hat{\mu}^{NT} y_{\cdot\cdot\cdot} \qquad (25)$$

Substituting Eq. (14) in Eq. (25) together with the algebraic simplification yields Eq. (24). This completes the proof. □

5 Sums of Squares for Incomplete Latin Square Design with One Missing experimental data

Proposition 3: In the $K \times K$ LSD with one missing experimental data, the sums of squares for the treatments, rows, and columns can be determined as

$$SS_{tr} = \frac{\sum_{All\ j}^{K} y_{\cdot j \cdot}^2}{K} + \frac{1}{(K-1)} \times \left[\frac{(y_{r\cdot\cdot} + y_{\cdot m\cdot} + y_{\cdot\cdot c} - y_{\cdots})^2}{K-2} - \frac{(y_{r\cdot\cdot} + y_{\cdot\cdot c} - y_{\cdots})^2}{K-1} + \frac{y_{\cdots}(2y_{\cdot m\cdot} - y_{\cdots})}{K} \right] \tag{26}$$

$$SS_{row} = \frac{\sum_{All\ i}^{K} y_{i\cdot\cdot}^2}{K} + \frac{1}{(K-1)} \times \left[\frac{(y_{r\cdot\cdot} + y_{\cdot m\cdot} + y_{\cdot\cdot c} - y_{\cdots})^2}{K-2} - \frac{(y_{\cdot m\cdot} + y_{\cdot\cdot c} - y_{\cdots})^2}{K-1} + \frac{y_{\cdots}(2y_{r\cdot\cdot} - y_{\cdots})}{K} \right] \tag{27}$$

$$SS_{column} = \frac{\sum_{All\ k}^{K} y_{\cdot\cdot k}^2}{K} + \frac{1}{(K-1)} \times \left[\frac{(y_{r\cdot\cdot} + y_{\cdot m\cdot} + y_{\cdot\cdot c} - y_{\cdots})^2}{K-2} - \frac{(y_{r\cdot\cdot} + y_{\cdot m\cdot} - y_{\cdots})^2}{K-1} + \frac{y_{\cdots}(2y_{\cdot\cdot k} - y_{\cdots})}{K} \right] \tag{28}$$

Proof. Based on Eq. (7), the treatment sum of squares (SS_{tr}) in Eq (26) can be derived by subtracting Eq. (24) in Preposition 2 from Eq. (12). The determinations of the row and column sums of squares are similarly carried out for SS_{tr} above. This completes the proof. □

An attracting illustration is given by an elongation experiment [25] which was laid out in a *5 × 5* LSD as shown in Table 4. There were five different versions of the stockings (treatments) by each of five investigators on five separate days.

Table 4. Elongation data [25]

Investigator (Row Variable)	**Day (Column Variable)**					Total
	1	**2**	**3**	**4**	**5**	
1	B = 22.1	A = 18.6	C = 23.0	E = 24.3	D = 17.1	$y_{1\cdot\cdot} = 105.1$
2	C = 23.5	D = 16.5	A = 18.7	B = 22.0	E = ---	$y_{2\cdot\cdot} = 80.7$
3	D = 17.4	E = 23.8	B = 22.8	C = 23.9	A = 20.0	$y_{3\cdot\cdot} = 107.9$

4	A = 20.3	B = 23.4	E = 25.9	D = 18.7	C = 24.2	$y_{4..} = 112.5$
5	E = 25.7	C = 24.8	D = 18.9	A = 20.6	B = 24.6	$y_{5..} = 114.6$
	$y_{..1} = 109$	$y_{..2} = 107.1$	$y_{..3} = 109.3$	$y_{..4} = 109.5$	$y_{..5} = 85.9$	$y_{...} = 520.8$
	$y_{.1.} = 98.2$	$y_{.2.} = 114.9$	$y_{.3.} = 119.4$	$y_{.4.} = 88.6$	$y_{.5.} = 99.7$	

The treatment, column, row and error sums of squares without bias can be easily calculated as presented in Table 5. In addition, the sum square of treatment using the missing plot technique is biased and cannot be used immediately in the ANOVA table, according to Ott and Longnecker [25].

Table 5. The analysis of the variance

Description	**Result**	**Remark**
Regression sum of squares for the full model	$R(\mu, \omega, \tau, \lambda) =$ 11,491.317	see Eq. (12)
Regression sum of squares for the reduced model	$R(\mu, \omega, \lambda) =$ 11,325.823	see Eq. (24)
Treatment sum of squares (Version)	$SS_{tr} =$ 165.4943 (with *df* = 4)	see Eq. (26)
Row sum of squares (Investigator)	$SS_{row} =$ 14.3688 (with *df* = 4)	see Eq. (27)
Column sum of squares (Day)	$SS_{column} =$ 0.9428 (with *df* = 4)	see Eq. (28)
Total sum of squares	$SS_{total} =$ 191.4000 (with *df* = 23)	see Eq. (5)
Error sum of squares	$SS_E =$ 13.2500 (with *df* = 11)	see Eq. (6)
Mean Square of Treatment	$MS_{tr} =$ 41.3736	see Table 3
Mean Square of Error	$MS_E =$ 0.1312	
F test-statistic for Treatment	*F test* = 315.35	see Table 3

6 Summary

An incomplete Latin square design (LSD) normally results in an unbalanced design, rendering the conventional sums of squares formulas invalid. Despite the prevalence of the missing-value techniques, the estimate of the missing experimental data is not experiment-based. Meanwhile, the existing exact approach failed to provide the simplified and straightforward mathematical formulas for the sub-variance calculation. This research has thus proposed the simplified exact approach based on the general linear model for solving the $K \times K$ LSD with one replicate and one missing experimental data. Under the proposed exact approach, the effects of the potential variable (the factor of primary interest) were determined by means of the regression sums of squares of the full and reduced-treatment models. In addition to the ease of computation, the treatment, row and column sums of squares are unbiased. More importantly, the mathematical expressions could be applied to the LSD with one missing experimental data of any order.

References

1. Montgomery, D.C.: Design and Analysis of Experiments. 7th edn. John Wiley & Sons, New York (2008)
2. Speed, T.P.: Introduction to Fisher (1926) The Arrangement of Field Experiments. In: Samuel Kotz S., Johnson N.L. (eds.), Breakthroughs in Statistics, vol. 2, pp. 71–81. Springer, New York (1992)
3. Mead, R., Gilmour, S. G., Mead, A.: Statistical Principles for the Design of Experiments: Applications to Real Experiments (Vol. 36). Cambridge University Press (2012)
4. Youden, W.J.: Use of Incomplete Block Replications in Estimating Tobacco-mosaic Virus. Contributions from Boyce Thompson Institute 9, vol. 1, pp. 41-48 (1937)
5. Yates, F.: Incomplete Randomized Blocks. Annals of Eugenics, vol. 7, pp.121-140 (1936)
6. Ai, M., Li, K., Liu, S., Lin, D.K.: Balanced Incomplete Latin Square Designs. Journal of Statistical Planning and Inference 9, vol. 143, pp. 1575-1582 (2013)
7. Allan, F. E., Wishart J.: A Method of Estimating the Yield of a Missing Plot in Field ExPerimental Work. The Journal of Agricultural Science 20, vol. 3, pp. 399-406 (1930)
8. Yates, F.: The Analysis of Replicated Experiments When the Field Results Are Incomplete. Empire Journal of Experimental Agriculture 2, vol. 1, pp. 129-142 (1933)
9. Sirikasemsuk, K.: A Review on Incomplete Latin Square Design of Any Order. In: Rusli N., Zaimi, W.M., Khazali, K.A., Masnan, M.J., Daud, W.S., Abdullah, N., Yahya, Z., Amin, N.A., Aziz N.H., Yusuf Y.N. (eds.) AIP Conference Proceedings 2016, vol. 1775, pp. 030022. AIP Publishing (2016)
10. Little, R.J.A., Rubin D.B.: Statistical Analysis with Missing Data. 2nd edn. John Wiley & Sons. (2002)
11. Coons, I.: The Analysis of Covariance as a Missing Plot Technique. Biometrics 3, vol. 13, pp. 387-405 (1957)
12. Cochran, W.G.: Analysis of Covariance: Its Nature and Uses. Biometrics 3, vol. 13, pp. 261-281 (1957)
13. Wilkinson, G. N.: Estimation of Missing Values for the Analysis of Incomplete Data. Biometrics 2, vol. 14, pp. 257-286 (1958)
14. Bartlett, M.S.: Some Examples of Statistical Methods of Research in Agriculture and Applied Biology. Supplement to the Journal of the Royal Statistical Society, pp. 137-183 (1937)
15. Kramer, C.Y., Glass, S.: Analysis of Variance of a Latin Square Design with Missing Observations. Applied Statistics, pp. 43-50 (1960)
16. Rubin, D.B.: A Non-iterative Algorithm for Least Squares Estimation of Missing Values in any Analysis of Variance Design. Applied Statistics, pp. 136-141 (1972)
17. Healy, M., Westmacott, M.: Missing Values in Experiments Analysed on Automatic Computers. Applied Statistics, pp. 203-206 (1956)
18. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (methodological), pp. 1-38 (1977)
19. Rubin, D.B.: Multiple Imputation for Non-response in Surveys. John Wiley & Sons (1987)
20. Zhang, H.: 25 New r-self-orthogonal Latin Squares. Discrete Mathematics, vol. 313, pp. 1746-1753 (2013)
21. Donovan D.M. Yazıcı E.Ş.: A Polynomial Embedding of Pairs of Orthogonal Partial Latin Squares. Journal of Combinatorial Theory, Series A, vol. 126, pp. 24-34 (2014)
22. Euler R.: On the Completability of Incomplete Latin Squares. European Journal of Combinatorics, vol. 31, pp. 535-552 (2010)
23. Casselgren C.J., Häggkvist, R.: Completing Partial Latin Squares with One Filled Row, Column and Symbol. Discrete Mathematics, vol. 313, pp. 1011-1017 (2013)
24. Sirikasemsuk, K.: One Missing Value Problem in Latin Square Design of Any Order: Regression Sum of Squares. In: 2016 Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems, pp. 142-147. IEEE Press, Japan (2016)
25. Ott, R.L., Longnecker, M.: An Introduction to Statistical Methods and Data Analysis. 6th edn. Cengage Learning. (2010)

Appendix

In this current research, the notations and their respective definitions are provided below:

y_{ijk}	the ijk^{th} observation taken under row *i*, column *k* and treatment *j*
i	index of rows (i = 1, 2, 3, …, K)
j	index of treatments (j = 1, 2, 3, …, K)
k	index of columns (k = 1, 2, 3, …, K)
K	the order of LSD
μ	the common effect or the overall mean of the observations
ω_i	the i^{th} row effects
τ_j	the j^{th} treatment effects
λ_k	the k^{th} column effects
ε_{ijk}	the normally distributed zero-mean random errors in the ijk^{th} observation
$\hat{\mu}$	the estimate of the parameter of μ
$\hat{\omega}_i$	the estimate of the parameter of the i^{th} row effect
$\hat{\tau}_j$	the estimate of the parameter of the j^{th} treatment effect
$\hat{\lambda}_k$	the estimate of the parameter of the k^{th} column effect
$y_{...}$	the grand total
$y_{..k}$	the k^{th} column total
$y_{.j.}$	the j^{th} treatment total
$y_{i..}$	the i^{th} row total

SS_{tr} the treatment sum of squares

SS_{row} the row sum of squares

SS_{column} the column sum of squares

SS_E the error sum of squares

SS_{total} the total sum of squares

$\hat{\mu}^{NT}$ the estimate of μ for the reduced model ignoring the treatment effect

$\hat{\omega}_i^{NT}$ the estimate of ω_i for the reduced model ignoring the treatment effect

$\hat{\lambda}_k^{NT}$ the estimate of λ_k for the reduced model ignoring the treatment effect

r index of the row in which the observation is missing

m index of the treatment (letter) in which the observation is missing

c index of the column in which the observation is missing

$\hat{\omega}_r$ the parameter estimate of the r^{th} row effect for the full effect model

$\hat{\tau}_m$ the parameter estimate of the m^{th} treatment effect for the full effect model

$\hat{\lambda}_c$ the parameter estimate of the c^{th} column effect for the full effect model

$\hat{\omega}_r^{NT}$ the estimate of ω_r for the reduced model ignoring the treatment effect

$\hat{\lambda}_c^{NT}$ the estimate of λ_c for the reduced model ignoring the treatment effect

$R(\mu, \omega, \tau, \lambda)$ the regression sum of squares for the full effect model of y_{ijk}

$R(\mu, \omega, \lambda)$ the regression sum of squares for the reduced-treatment effect model of y_{ijk}

Public interest statement

On the Fisher's principles, a classical experimental design (e.g., one-way ANOVA, Latin square design, 2-level factorial design, factional factorial design, and so on) is a powerful methodology in order to explain causal mechanisms between independent variables and response variable by means of the identification of variation of data. Latin square design is of great use for analyzing one potential variable and two block variables. One missing experimental data could, however, pose significant challenges to the analysis. In this research, an incomplete Latin square design of any order with one missing experimental data was of the exact approach based on the general linear model. Due to the lack of ready-made formula, this research paper has thus proposed the explicit and mathematical formulae for the treatment sum of squares for ease of comparisons of mean squares, along with an F-test.

About the author

Kittiwat Sirikasemsuk is an Assistant Professor in Industrial Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand. He received Doctor of Philosophy (Ph.D.) from Industrial Systems Engineering, Asian Institute of Technology, Thailand, in 2013. He has extensive experiences in lean manufacturing and various quality engineering techniques. His research activities cover a wide range of area in: design of experiments, supply chain design, measures of bullwhip effect in supply chains, quality engineering, lean manufacturing, etc.

Kanogkan Leerojanaprapa is a lecturer in Statistics department, King Mongkut's Institute of Technology Ladkrabang, Thailand. She holds a Ph.D. in Management Science since 2014 from University of Strathclyde, UK. Her research area includes quality control, risk analysis, supply chain management, bayesian network, and applications of statistical models. She has published articles in various peer reviewed international journals and conferences.