cogent engineering

PRODUCTION & MANUFACTURING | RESEARCH ARTICLE

One missing value problem in Latin square design of any order: Exact analysis of variance

Kittiwat Sirikasemsuk[1]* and Kanogkan Leerojanaprapa[2]

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

Abstract: This research proposes a simplified exact approach based on the general linear model for solving the $K \times K$ Latin square design (LSD) 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 LSD with one missing value of any order. Moreover, the treatment, row and column sums of squares are unbiased.

Subjects: Mathematical Modeling; Mathematical Statistics; Statistical Computing; Statistics & Computing; Quality Control & Reliability

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 (Montgomery, 2008). 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

ABOUT THE AUTHORS

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 (PhD) 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 PhD 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.

PUBLIC INTEREST STATEMENT

On the Fisher's principles, a classical experimental design (e.g. one-way ANOVA, Latin square design (LSD), 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. LSD 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 LSD 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.

resembles the desired nominal value. In fact, Fisher's classical DOE is a form of statistical hypothesis testing under the analysis of variance (ANOVA) (Speed, 1992). Meanwhile, ANOVA is defined as a collection of statistical procedures to compare the between-group variation with the within-group variation (Montgomery, 2008).

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 (Mead, Gilmour, & Mead, 2012) 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 Youden (1937), 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 (1937), Yates (1936), and Ai, Li, Liu, and Lin (2013), 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 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 ANOVA in the incomplete-observation experimental design

The work of Allan and Wishart (1930) 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 Yates (1933) and Sirikasemsuk (2016a), 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 (Little & Rubin, 2002). In Coons (1957), Cochran (1957) and Wilkinson (1958), 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 (1937).

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.

Many recent research studies considered aspects of combinatorics, examples of which were the studies on the construction of the orthogonal Latin squares by Zhang (2013) and Donovan and Şule Yazıcı (2014); and the studies on the completability of the incomplete Latin squares from the partial Latin squares by Euler (2010) and Casselgren and Häggkvist (2013).

cogent ·· engineering

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 (1930)
	Differentiating the error sum of squares to each missing value (when only one observation is missing)	Yates (1933)
	General method for estimating several missing values in Latin square design	Kramer and Glass (1960)
	Non-iterative Rubin method	Rubin (1972)
Missing plot technique with iterative method	Iterative Yates method (based on the work of Allan and Wishart (1930) when more than one observations are missing)	Yates (1933)
	Healy-Westmacott method based on regression imputation	Healy and Westmacott (1956)
Exact approach with general linear model	General regression significance test	Montgomery (2008)
Analysis of covariance (ANCOVA) technique	A combination of regression analysis and ANOVA consisting of the covariate	Coons (1957), Cochran (1957), and Wilkinson (1958)
Expectation maximization algorithm (EM Algorithm)	Iterative method with maximum likelihood estimation	Dempster, Laird, and Rubin (1977)
Multiple imputation (MI) method	A combination of raw maximum likelihood and EM method	Rubin (1987)

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 (Montgomery, 2008; Sirikasemsuk, 2016a). 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 $K \times K$ LSD

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

$$y_{ijk} = \mu + \omega_i + \tau_j + \lambda_k + \varepsilon_{ijk} \tag{1}$$

where ε_{ijk} is independently, identically and normally distributed, i.e. $\varepsilon_{ijk} \sim N(0, \sigma^2)$.

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

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

❊ cogent ·· engineering

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

Blocking variable 1	Blocking variable 2						
	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}		

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

$$SS_{column} = \sum_{k=1}^{K} \left(\hat{\lambda}_k\right)^2 = \sum_{k=1}^{K} \left(\bar{y}_{..k} - \bar{y}_{...}\right)^2 \tag{4}$$

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

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

3. Incomplete LSD and regression sum of squares under the full model

For the missing-data LSD, the sums of squares in Equations (2)–(4) are invalid. The general regression significance test could instead be applied to the incomplete LSD for ANOVA. According to Montgomery (2008), the computational formulas for the sums of squares of treatments, rows, columns and errors could respectively be expressed as

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

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

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

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

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

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

In Sirikasemsuk (2016b), the estimated values of all parameters (Equation (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

$$R(\mu, \omega, \tau, \lambda) = \frac{\sum_{All\ i}^{K} y_{i..}^2 + \sum_{All\ j}^{K} y_{.j.}^2 + \sum_{All\ k}^{K} y_{..k}^2}{K} + \frac{(1-K)\left(2y_{...}^2 - y_{sum\_m}^2\right) + \left(y_{sum\_m} - 2y_{...}\right)^2}{K(K-1)(K-2)} \qquad (12)$$

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

To find the treatment sum of squares (see Equation (7)), it is assumed that the treatment effects (τ_j) are not considered in Equation (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 Equation (13).

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

The estimated model parameters, i.e. $\hat{\mu}^{NT}, \hat{\omega}_i^{NT}$ 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 Equation (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 Equations (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

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

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

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

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

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

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

$$\omega_i: K\hat{\mu}^{NT} + K\hat{\omega}_i^{NT} + \sum_{k=1}^{K} \hat{\lambda}_k^{NT} = y_{i..} \tag{20}$$

$$\lambda_k: K\hat{\mu}^{NT} + \sum_{i=1}^{K} \hat{\omega}_i^{NT} + K\hat{\lambda}_k^{NT} = y_{..k} \tag{21}$$

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

$$\hat{\omega}_i^{NT} = \frac{y_{i..}}{K} - \hat{\mu}^{NT} \tag{22}$$

$$\hat{\lambda}_k^{NT} = \frac{y_{..k}}{K} - \hat{\mu}^{NT} \tag{23}$$

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

$$R(\mu, \omega, \lambda) = \frac{\sum_{All\ i=1}^{K} y_{i..}^2 + \sum_{All\ k=1}^{K} y_{..k}^2}{K} + \frac{(y_{r..} + y_{..c} - y_{...})\left[K(y_{r..} + y_{..c}) + (K-2)y_{...}\right]}{K(K-1)^2} \tag{24}$$

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 (2016b) and is presented as below.

Substituting Equations (15), (16), (22) and (23) into Equation (13), we obtain

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

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

5. Sums of squares for incomplete LSD 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

$$SS_{tr} = \frac{\sum\limits_{All\ j}^{K} y_{.j.}^2}{K} + \frac{1}{(K-1)} \times \left[\frac{(y_{r..} + y_{.m.} + y_{..c} - y_{...})^2}{K-2} - \frac{(y_{r..} + y_{..c} - y_{...})^2}{K-1} + \frac{y_{...}(2y_{.m.} - y_{...})}{K} \right] \qquad (26)$$

$$SS_{row} = \frac{\sum\limits_{All\ i}^{K} y_{i..}^2}{K} + \frac{1}{(K-1)} \times \left[\frac{(y_{r..} + y_{.m.} + y_{..c} - y_{...})^2}{K-2} - \frac{(y_{.m.} + y_{..c} - y_{...})^2}{K-1} + \frac{y_{...}(2y_{r..} - y_{...})}{K} \right] \qquad (27)$$

$$SS_{column} = \frac{\sum\limits_{All\ k}^{K} y_{..k}^2}{K} + \frac{1}{(K-1)} \times \left[\frac{(y_{r..} + y_{.m.} + y_{..c} - y_{...})^2}{K-2} - \frac{(y_{r..} + y_{.m.} - y_{...})^2}{K-1} + \frac{y_{...}(2y_{..k} - y_{...})}{K} \right] \qquad (28)$$

Proof. Based on Equation (7), the treatment sum of squares (SS_{tr}) in Eq (26) can be derived by subtracting Equation (24) in Preposition 2 from Equation (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 (Ott & Longnecker, 2010) 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.

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 (2010).

Table 4. Elongation data						
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	$y_{1..}$ = 105.1
2	C = 23.5	D = 16.5	A = 18.7	B = 22.0	E = –	$y_{2..}$ = 80.7
3	D = 17.4	E = 23.8	B = 22.8	C = 23.9	A = 20.0	$y_{3..}$ = 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	

Source: Ott and Longnecker (2010).

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 Equation (12)
Regression sum of squares for the reduced model	$R(\mu, \omega, \lambda) = 11{,}325.823$	See Equation (24)
Treatment sum of squares (version)	$SS_{tr} = 165.4943$ (with $df = 4$)	See Equation (26)
Row sum of squares (investigator)	$SS_{row} = 14.3688$ (with $df = 4$)	See Equation (27)
Column sum of squares (day)	$SS_{column} = 0.9428$ (with $df = 4$)	See Equation (28)
Total sum of squares	$SS_{total} = 191.4000$ (with $df = 23$)	See Equation (5)
Error sum of squares	$SS_{E} = 13.2500$ (with $df = 11$)	See Equation (6)
Mean square of treatment	$MS_{tr} = 41.3736$	See Table 3
Mean square of error	$MS_{E} = 0.1312$	See Table 3
F test-statistic for treatment	F test $= 315.35$	See Table 3

6. Summary

An incomplete 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.

Author details
Kittiwat Sirikasemsuk[1]
E-mail: kittiwat.sirikasemsuk@gmail.com
ORCID ID: http://orcid.org/0000-0001-8957-0947
Kanogkan Leerojanaprapa[2]
[1] Faculty of Engineering, Department of Industrial Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.
[2] Faculty of Science, Statistics Department, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.

References
Ai, M., Li, K., Liu, S., & Lin, D. K. (2013). Balanced incomplete Latin square designs. *Journal of Statistical Planning and Inference, 143*(9), 1575–1582. https://doi.org/10.1016/j.jspi.2013.05.001
Allan, F. E., & Wishart, J. (1930). A method of estimating the yield of a missing plot in field experimental work. *The Journal of Agricultural Science, 20*(3), 399–406. https://doi.org/10.1017/S0021859600006912
Bartlett, M. S. (1937). Some examples of statistical methods of research in agriculture and applied biology. *Supplement to the Journal of the Royal Statistical Society, 4*, 137–183. https://doi.org/10.2307/2983644
Casselgren, C. J., & Häggkvist, R. (2013). Completing partial Latin squares with one filled row. *Column and Symbol. Discrete Mathematics, 313*, 1011–1017. https://doi.org/10.1016/j.disc.2013.01.019
Cochran, W. G. (1957). Analysis of covariance: Its nature and uses. *Biometrics, 13*(3), 261–281. https://doi.org/10.2307/2527916
Coons, I. (1957). The analysis of covariance as a missing plot technique. *Biometrics, 13*(3), 387–405. https://doi.org/10.2307/2527922
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological), 39*, 1–38.
Donovan, D. M., & Şule Yazıcı, E. (2014). A polynomial embedding of pairs of orthogonal partial Latin squares. *Journal of Combinatorial Theory, Series A, 126*, 24–34. https://doi.org/10.1016/j.jcta.2014.04.003
Euler, R. (2010). On the completability of incomplete Latin squares. *European Journal of Combinatorics, 31*, 535–552. https://doi.org/10.1016/j.ejc.2009.03.036
Healy, M., & Westmacott, M. (1956). Missing values in experiments analysed on automatic computers. *Applied Statistics, 5*, 203–206. https://doi.org/10.2307/2985421
Kramer, C. Y., & Glass, S. (1960). Analysis of variance of a Latin square design with missing observations. *Applied Statistics, 9*, 43–50. https://doi.org/10.2307/2985758
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119013563

Mead, R., Gilmour, S. G., & Mead, A. (2012). *Statistical principles for the design of experiments: Applications to real experiments*, Vol. 36. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139020879

Montgomery, D. C. (2008). *Design and analysis of experiments* (7th ed.). New York, NY: John Wiley & Sons.

Ott, R. L., & Longnecker, M. (2010). *An introduction to statistical methods and data analysis* (6th ed.). Boston: Cengage Learning.

Rubin, D. B. (1972). A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Applied Statistics,* 136–141. https://doi.org/10.2307/2346485

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*. Hoboken: John Wiley & Sons. https://doi.org/10.1002/SERIES1345

Sirikasemsuk, K. (2016a). A review on incomplete latin square design of any order. In N. Rusli, W. M. Zaimi, K. A. Khazali, M. J. Masnan, W. S. Daud, N. Abdullah, ..., Y. N. Yusuf (Eds.), *AIP Conference Proceedings 2016*, (vol. 1775, pp. 030022). Melville: AIP Publishing.

Sirikasemsuk, K. (2016b). 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). Japan: IEEE Press.

Speed, T. P. (1992). Introduction to fisher (1926) the arrangement of field experiments. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in statistics*, Vol. 2 (pp. 71–81). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4380-9

Wilkinson, G. N. (1958). Estimation of missing values for the analysis of incomplete data. *Biometrics, 14*(2), 257–286. https://doi.org/10.2307/2527789

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Empire Journal of, Experimental Agriculture, 2*(1), 129–142.

Yates, F. (1936). *Incomplete randomized blocks. Annals of eugenics, 7*, 121–140. https://doi.org/10.1111/j.1469-1809.1936.tb02134.x

Youden, W. J. (1937). Use of incomplete block replications in estimating tobacco-mosaic virus. *Contributions from Boyce Thompson Institute, 9*(1), 41–48.

Zhang, H. (2013). 25 new r -self-orthogonal Latin squares. *Discrete Mathematics, 313*, 1746–1753. https://doi.org/10.1016/j.disc.2013.04.021

cogent • engineering

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 ith row effects
τ_j	the jth treatment effects
λ_k	the kth 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 ith row effect
$\hat{\tau}_j$	the estimate of the parameter of the jth treatment effect
$\hat{\lambda}_k$	the estimate of the parameter of the kth column effect
$y_{...}$	the grand total
$y_{..k}$	the kth column total
$y_{.j.}$	the jth treatment total
$y_{i..}$	the ith 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 rth row effect for the full effect model
$\hat{\tau}_m$	the parameter estimate of the mth treatment effect for the full effect model
$\hat{\lambda}_c$	the parameter estimate of the cth 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}

cogent •• engineering

Cogent Engineering (ISSN: 2331-1916) is published by Cogent OA, part of Taylor & Francis Group.

Publishing with Cogent OA ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a Cogent OA journal at www.CogentOA.com