Pak. J. Statist.
2020 Vol. 36(3), 207-224

A NON-ITERATIVE LEAST SQUARES ESTIMATION OF MISSING DATA IN RECTANGULAR LATTICE DESIGNS

Abimibola Victoria Oladugba[1] and **Emmanuel Ogochukwu Ossai**[2]
Department of Statistics, University of Nigeria, Nsukka, Nigeria
Email: [1]abimibola.oladugba@unn.edu.ng
[2]emmanuel.ossai.pg77374@unn.edu.ng

ABSTRACT

Missing data are inevitable problem in almost every research such as experimental and observational researches. In this paper, we proposed a non-iterative least square formulae for estimating missing data in rectangular lattice designs (simple and triple) without repetitions of the basic design using the intra-block information when one or more observations are missing. A numerical example was used to show the applicability of the proposed formulae in terms of estimated values, standard errors and p-value. Results showed that the estimated values were significantly approximate to the complete data which indicate the effectiveness of the proposed formulae for estimating missing data in rectangular lattice designs without repetitions.

KEY WORDS

Rectangular lattice designs, treatments, bias, repetition, missing data, non-iterative least square estimate.

1. INTRODUCTION

Rectangular lattice designs are set of resolvable incomplete block designs in which the number of treatments is the product of two consecutive integers, and the number of replications of each treatment is either two or three which could be with or without repetition of the basic designs, Harshbarger (1946, 1947, and 1949) and Cochran and Cox (1957). Rectangular lattice design is an experimental design involving $t = k(k + 1)$ treatments in $b = (k + 1)$ incomplete blocks each containing k units arranged into r complete replications denoted by X , Y , and Z. Rectangular lattice design can either be simple or triple. It is said to be simple when the first two replications (say X and Y) from any plan are used and triple when three replications (say X, Y and Z) of the plan are used, Singh (2019a, b). A rectangular lattice design is said to be with repetition when the basic design with r replications is repeated p times to give $n = pr$ replications, otherwise, it is without repetition, Ossai and Oladugba (2018). The importance of these designs is that they form a useful addition to the square lattice designs since the allowable number of treatments fall about midway between the allowable numbers for square lattices, Cochran and Cox (1957). Rectangular lattice designs are often used in large experiments involving several observations, it is not easy to ensure that all the observations are accurately made. Even with careful management of the experiment, there is always a chance that mistakes or accidents leading to missing data will affect a few of the observations, and inferences from the experiments.

In experimental or observational study, it frequently happens that one or more experimental units are missing from the dataset, Ossai and Oladugba (2018), Geeta et al., (2017), Coertjens et al., (2017), Takahashi (2017), Sullivan et al., (2017), Meiller et al., (2017), Schmitt et al., (2015) and Nuryazmin et al., (2015). The uncontrollable existence of missing data in designed experiment has always been a problem of interest to researchers, Ossai and Oladugba (2018). Missing data occur for various reasons, such as; incorrect application of treatment, destruction of experimental units, non-response and inability to measure certain attributes which results in loss of information, non-applicability of the standard form of analysis and introduction of a new problem, Little and Rubin (2002), Azadeh et al., (2013), Schmitt et al., (2015) and Mieller et al., (2017). It is necessary to deal with missing values in a dataset since data with missing values can cause a significant problem in the analysis and precision of results, Nuryazmin et al., (2015) and Alrweili et al., (2019). This is the motivation behind this current study, to propose a simplified and an efficient procedure for estimating one or more number of missing data in rectangular lattice design, since missing data tend to be common with them.

2. MISSING DATA TECHNIQUES

The easiest way of dealing with missing data is to delete all incomplete cases and continue the analysis only with the complete cases, Azadeh et al., (2013), Geeta et al., (2017), Coertjens et al., (2017) or by listwise deletion method which simply removing the missing data and uses the remaining data set for analysis, Nuryazmin et al., (2015). Though these techniques can be convenient in simplifying the problem at hand, Little and Rubin (2002), but can generate serious bias (especially when the mechanism of the missing data is not missing at random), loss of valuable information therefore leading to inefficiency and inaccurate cost estimation models, particularly when the number of such cases is large as compared to the sample size and/or when there is a specific reason for missing data relevant to the study, Geeta (2017), Gad and Ahmed (2006), Takahashi (2017), Sullivan et al., (2017) and Ossai and Oladugba (2018). Another common method is by using the average values for imputation of the missing values, Allison (2001). This is the easiest imputation method which imputes the mean value of each variable on the respective missing variables as an estimate of the missing value. This method can lead to a problem of bias and large errors in the covariance matrix thus affecting the performance of the statistical modeling. Other approaches of imputation techniques that can be used are the hot deck imputation (which imply the nearest neighbor method) and the imputation which is based on the least squares and maximum likelihood, Nuryazmin et al., (2015). Missing data have been treated adequately in various real world datasets involving experimental designs such as randomized block design, incomplete block design, Latin square design, Graeco-Latin square design, F-square design, cross-over design, split-plot design and nested-factorial design, Bhatra and Dharamyadav (2013), Subramani (1991a,b, 1994), Ahmed (2016) and Okereke et al., (2018) and rectangular lattice designs with repetition, Ossai and Oladugba (2018).

The purpose of this paper is to derive non-iterative least square formulae for estimating several missing data in rectangular lattice designs (simple and triple rectangular lattice designs) without repetition of the basic design using the intra-block information. The non-iterative least square technique minimizes the intra-block error sum of squares with respect to the missing data, and solve the resultant to obtain an estimate for the missing data. A numerical example was used to show the applicability of these formulae.

3. METHODOLOGY

3.1 Model for Rectangular Lattice Designs without Repetition

The statistical linear model for rectangular lattice design without repetition for $t = k (k + 1)$ treatments in $b = k + 1$ incomplete blocks each containing k units arranged into r complete replications is given by (1) as

$$Y_{ij(l)} = \mu + \pi_j + \beta_{l(j)} + \tau_i + \varepsilon_{ij(l)} \; ; i = 1,\ 2,\ \ldots ,\ k(k+1)\ , j = 1,\ 2,\ \ldots ,\ r\ , l = 1,\ 2,\ \ldots ,\ k+1 \quad (1)$$

where $Y_{ij(l)}$ = denotes the response value of the i^{th} treatment in the l^{th} block within j^{th} replication; μ, π_j, $\beta_{l(j)}$ and τ_i represent the effect of the mean, the replicate, the incomplete block and treatment, respectively; $\varepsilon_{ij(l)}$ is the intra-block disturbance, assumed to be normally distributed with mean zero (0) and variance, σ^2.

Table 1 shows the analysis of variance (ANOVA) table associated with (1), where R_j is the sum of the yields of the treatments for j^{th} replicate; τ_i is the sum of the yields from all replicates of the treatments; C_{jl} is the total (over all replicates) of all treatments in the block subtracted from rB_{jl} : r is the number of replications; B_{jl} is the sum of the k units in the l^{th} block of the j^{th} replicate; R_c is the sum of the C_{jl} in the j^{th} replicate; S is the sum of the same peer in the same row; G is the grand total; k is the block size; ε_b is the block error and ε_e is the intra-block error.

Table 1
ANOVA table for *k* (*k* + 1) Rectangular Lattice Design without Repetition

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Replicates (SS_R)	r -1	$\frac{\sum R_j^2}{k(k+1)} - \frac{G^2}{rk(k+1)}$	
Treatments (SS_T) (Unadj.)	$k^2 + k - 1$	$\frac{\sum \tau_i^2}{r} - \frac{G^2}{rk(k+1)}$	
Blocks within (rep.) (SS_B) Adj.	rk	$\frac{\sum C_{jk}^2}{r(rk-k-1)} - \frac{\sum R_c^2}{r(k+1)(rk-k-1)} - \frac{\sum S_k^2}{r(r-1)(k+1)(rk-k-1)}$	ε_b (error)
Intra-block error (SS_E)	$(r - 1)$ $(k^2 - 1) - k$	By Subtraction	ε_e (error)
Total (SS_{TOTAL})	$rk^2 + rk - 1$	$\sum Y_{ijk}^2 - \frac{G^2}{rk(k+1)}$	

3.2 Estimation of Missing data in Rectangular Lattice Designs for Single Observation

Suppose that yield from the plot in replicate *j*, block *l*, is missing and this plot receive treatment *i*. Denote the missing plot by *u*. Let R_j, τ_i, C_{jl}, R_c, S and G retain their meaning in (3.1), so that R'_j is the sum of the observations of all the treatments in replicate *j*; τ'_i is the sum of the observations from all replicates of treatment *i*; C'_{jl} is the total (over replicate *j*) of all treatments in block *l* subtracted from rB'_{jl}; R'_c is the sum of the C_{jl} in replicate *j*; S'_l is the sum of the observations in the same row and G' is the unknown grand total. The least square procedure is to minimize the intra-block error sum of squares (SS_E) with respect to the missing plot *u* and perform the analysis of variance. This is shown in Table 2.

Table 2
Modified ANOVA table

Source of Variation	Sum of Squares
Replicates (SS_R)	$\frac{(R'_j+u)^2}{k(k+1)}+\frac{\sum R_j^2}{k(k+1)}-\frac{(G'+u)^2}{rk(k+1)}$
Treatments (SS_T) (Unadj.)	$\frac{(\tau'_i+u)^2}{r}+\frac{\sum \tau_i^2}{r}-\frac{(G'+u)^2}{rk(k+1)}$
Blocks within (rep.) (SS_B) Adj.	$\frac{(C'_{jl}+u)^2}{r(rk-k-1)}+\frac{\sum C_{jl}^2}{r(rk-k-1)}-\frac{(R'_c+u)^2}{r(k+1)(rk-k-1)}-\frac{\sum R_c^2}{r(k+1)(rk-k-1)}-\frac{(S'_l+u)^2}{r(r-1)(k+1)(rk-k-1)}+\frac{\sum S_l^2}{r(r-1)(k+1)(rk-k-1)}$
Intra-block error (SS_E)	By Subtraction
Total (SS_{TOTAL})	$u^2+\sum Y_{ijk}^2-\frac{(G'+u)^2}{rk(k+1)}$

From Table 2, SS_E given by (2) is obtained by subtraction as

$$SS_E = u^2+\frac{(G'+u)^2}{rk(k+1)}-\frac{(R'_j+u)^2}{k(k+1)}-\frac{(\tau'_i+u)^2}{r}-\frac{(C'_{jl}+u)^2}{r(rk-k-1)}$$
$$-\frac{(R'_c+u)^2}{r(k+1)(rk-k-1)}-\frac{(S'_l+u)^2}{r(r-1)(k+1)(rk-k-1)}+Q \qquad (2)$$

and Q is all terms not containing *u*.

Differentiating SS_E in (2) with respect to *u* and, setting the derivative equal to zero, we have;

$$\frac{\partial(SS_E)}{\partial u} = 2u + \frac{2(G'+u)}{rk(k+1)} - \frac{2(R'_j+u)}{k(k+1)} - \frac{2(\tau'_i+u)}{r} - \frac{2(C'_{jl}+u)}{r(rk-k-1)} + \frac{2(R'_c+u)}{r(k+1)(rk-k-1)} + \frac{2(S'_l+u)}{r(k+1)(rk-k-1)} = 0 \quad (3)$$

Then solving (3) for u,

$$u = \frac{(rk-k-1)(r-1)[rR'_j + k(k+1)\tau'_i - G'] + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[(r-1)^2(k^2+k-1)-k]} \quad (4)$$

Equation (4) gives the non-iterative least square estimate for estimating single missing observation in simple rectangular lattice (SRL) and triple rectangular lattice (TRL) designs, where $r = 2$ and 3 respectively for SRL and TRL designs.

3.3 Estimation of Missing Data in Rectangular Lattice Designs for Two Observations

There are several ways in which two values can be missing in SRL and TRL designs without repetition such as: different treatments missing in different replicates, but same block; different treatments missing in same replicate, but same block; different treatments missing in same replicate, but different blocks; and same treatments missing in different replicates, but different blocks.

Suppose u and v are the two values missing in replicate X and Y respectively, given any of the cases below, we differentiate the sum of squares errors with respect to u and v and equate to zero to give a system of simultaneous equations for estimating the two missing values. We define Q_1, Q_2, Q_3 and Q_4 as all terms not containing u and v for the 4 different cases considered. The missing u and v can occur either in replicate I (X) or replicate II (Y) or replicate III (Z) respectively. Thus the derived formulae for the basic design can be extended to any replicate depending on where the missing values occur. The estimate obtained for u and v in the different cases looks similar but they are different in terms of R'_j, τ'_i, C'_{jl}, R'_c C_{jl}, S'_l and G'.

Case 1: Different treatments missing in different replicates, but same block

Let u and v be missing observations in rectangular lattice designs given this case, and $A = k(k + 1)$, $B = rk(k + 1)$, $C = r(rk - k - 1)$, $D = r(k + 1)(rk - k - 1)$, $E = r(r - 1)(k + 1)(rk - k - 1)$. The intra-block sum of squares error (SS_E) for this case given by (5), is obtained by observing the missing values u and v in the sources of variation of Table 2 as

$$SS_E = u^2 + v^2 + \left(\frac{(G'+u+v)^2}{B}\right) - \left(\frac{(R'_j+u)^2 + (R'_j+v)^2}{A}\right) - \left(\frac{(\tau'_i+u)^2 + (\tau'_i+v)^2}{r}\right) - \left(\frac{(C'_{jl}+u)^2 + (C'_{jl}+v)^2}{C}\right) + \left(\frac{(R'_c+u)^2 + (R'_c+v)^2}{D}\right) + \left(\frac{(S'_l+u+v)^2}{E}\right) + Q_1 \quad (5)$$

Differentiating (5) with respect to the missing plots, u and v, and equating to zero we have;

$$2u+\frac{2\left(G'+u+v\right)}{B}-\frac{2\left(R_j'+u\right)}{A}-\frac{2\left(\tau_i'+u\right)}{r}-\frac{2\left(C_{jl}'+u\right)}{C}+\frac{2\left(R_c'+u\right)}{D}+\frac{2\left(S_l'+u+v\right)}{E}=0 \quad (6)$$

$$2v+\frac{2\left(G'+u+v\right)}{B}-\frac{2\left(R_j'+v\right)}{A}-\frac{2\left(\tau_i'+v\right)}{r}-\frac{2\left(C_{jl}'+v\right)}{C}+\frac{2\left(R_c'+v\right)}{D}+\frac{2\left(S_l'+u+v\right)}{E}=0 \quad (7)$$

Solving simultaneously (6) and (7) to obtain estimates of the missing data, we have

$$u=\frac{HR_j'+I\tau_i'-(r-1)(rk-k-1)G'+k[(r-1)(k+1)C_{jl}'-(r-1)R_c'-S_l']}{k[k(r-1)(k(r-1)^2+(r^2-3r+1))-r(r^2-3r+3)]+(r^2-3r+2)} \quad (8)$$

$$v=\frac{HR_j'+I\tau_i'-(r-1)(rk-k-1)G'+k[(r-1)(k+1)C_{jl}'-(r-1)R_c'-S_l']}{k[k(r-1)(k(r-1)^2+(r^2-3r+1))-r(r^2-3r+3)]+(r^2-3r+2)} \quad (9)$$

where $H=r+kr-r^2-2kr^2+kr^3$ and $I=k(k+1)(k(r-1)-1)(r-1)$.

Equations (8) and (9) give the non-iterative least square estimates for estimating two missing observations in SRL and TRL designs when two different treatments are missing in different replicates, but same block.

Generally, for p missing values given case 1, we have,

$$u=\frac{HR_j'+I\tau_i'-(r-1)(rk-k-1)G'+k[(r-1)(k+1)C_{jl}'-(r-1)R_c'-S_l']}{k[k(r-1)(k(r-1)^2+(r^2-3r+1))-r(r^2-3r+3)]+(r^2-3r+2)}$$

$$v=\frac{HR_j'+I\tau_i'-(r-1)(rk-k-1)G'+k[(r-1)(k+1)C_{jl}'-(r-1)R_c'-S_l']}{k[k(r-1)(k(r-1)^2+(r^2-3r+1))-r(r^2-3r+3)]+(r^2-3r+2)}$$

$$\vdots \qquad\qquad \vdots$$

$$p=\frac{HR_j'+I\tau_i'-(r-1)(rk-k-1)G'+k[(r-1)(k+1)C_{jl}'-(r-1)R_c'-S_l']}{k[k(r-1)(k(r-1)^2+(r^2-3r+1))-r(r^2-3r+3)]+(r^2-3r+2)} \quad (10)$$

Case 2: Different treatments missing in same replicate, but same block

Similarly, as in case 1, the SS_E for this case is given by;

$$SS_E = u^2 + v^2 + \left(\frac{\left(G' + u + v\right)^2}{B} \right) - \left(\frac{\left(R'_j + u + v\right)^2}{A} \right) - \left(\frac{\left(\tau'_i + u\right)^2 + \left(\tau'_i + v\right)^2}{r} \right)$$
$$- \left(\frac{\left(C'_{jl} + u + v\right)^2}{C} \right) + \left(\frac{\left(R'_c + u + v\right)^2}{D} \right) + \left(\frac{\left(S'_l + u + v\right)^2}{E} \right) + Q_2 \qquad (11)$$

Differentiating (11) with respect to the missing plots, u and v, and equating to zero we have;

Solving simultaneously (12) and (13) to obtain estimates of the missing data, we have,

$$2u + \frac{2\left(G' + u + v\right)}{B} - \frac{2\left(R'_j + u + v\right)}{A} - \frac{2\left(\tau'_i + u\right)}{r} - \frac{\left(C'_{jl} + u + v\right)}{C}$$
$$+ \frac{2\left(R'_c + u + v\right)}{D} + \frac{2\left(S'_l + u + v\right)}{E} = 0 \qquad (12)$$

$$2v + \frac{2\left(G' + u + v\right)}{B} - \frac{2\left(R'_j + u + v\right)}{A} - \frac{2\left(\tau'_i + v\right)}{r} - \frac{\left(C'_{jl} + u + v\right)}{C}$$
$$+ \frac{2\left(R'_c + u + v\right)}{D} + \frac{2\left(S'_l + u + v\right)}{E} = 0 \qquad (13)$$

$$u = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[(r-1)^2(k^2-2) + k(r^2-2r-1)]} \qquad (14)$$

$$v = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[(r-1)^2(k^2-2) + k(r^2-2r-1)]} \qquad (15)$$

where H and I are as defined in Case 1.

Equations (14) and (15) give the non-iterative least square estimates for estimating two missing observations SRL and TRL designs when two different treatments are missing in same replicate, but same block.

Generally, for p missing values given case 2, we have,

$$u = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[(r-1)^2(k^2-2) + k(r^2-2r-1)]}$$
$$v = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[(r-1)^2(k^2-2) + k(r^2-2r-1)]}$$
$$\vdots \qquad\qquad \vdots$$
$$p = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[(r-1)^2(k^2-2) + k(r^2-2r-1)]} \qquad (16)$$

Case 3: Different treatments missing in same replicate, but different blocks

The intra-block sum of squares error (SS_E) for this case is;

$$SS_E = u^2 + v^2 + \left(\frac{(G' + u + v)^2}{B}\right) - \left(\frac{(R'_j + u + v)^2}{A}\right) - \left(\frac{(\tau'_i + u)^2 + (\tau'_i + v)^2}{r}\right)$$
$$- \left(\frac{(C'_{jl} + u)^2 + (C'_{jl} + v)^2}{C}\right) + \left(\frac{(R'_c + u + v)^2}{D}\right) + \left(\frac{(S'_l + u)^2 + (S'_l + v)^2}{E}\right) + Q_3 \quad (17)$$

Differentiating (17) with respect to the missing plots, *u* and *v*, and equating to zero we have;

$$2u + \frac{2(G' + u + v)}{B} - \frac{2(R'_j + u + v)}{A} - \frac{2(\tau'_i + u)}{r} - \frac{2(C'_{jl} + u)}{C} + \frac{2(R'_c + u + v)}{D} + \frac{2(S'_l + u)}{E} = 0 \quad (18)$$

$$2v + \frac{2(G' + u + v)}{B} - \frac{2(R'_j + u + v)}{A} - \frac{2(\tau'_i + v)}{r} - \frac{2(C'_{jl} + v)}{C} + \frac{2(R'_c + u + v)}{D} + \frac{2(S'_l + v)}{E} = 0 \quad (19)$$

Solving simultaneously (18) and (19) to obtain estimates of the missing data, we have,

$$u = \frac{HR'_j + I\tau'_i - (r-1)(rk - k - 1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(r-1)[(r-1)(rk^3 - k^3 + 2) + k^2(r^2 - 3r + 1)] + k(5r^2 - 2r^3 - 3r + 1)]} \quad (20)$$

$$v = \frac{HR'_j + I\tau'_i - (r-1)(rk - k - 1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(r-1)[(r-1)(rk^3 - k^3 + 2) + k^2(r^2 - 3r + 1)] + k(5r^2 - 2r^3 - 3r + 1)]} \quad (21)$$

where *H* and *I* are as defined in Case 1.

Equations (20) and (21) give the non-iterative least square estimates for estimating two missing observations in SRL and TRL designs when two different treatments are missing in same replicate, but different blocks.

Generally, for *p* missing values given case 3, we have,

$$u = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(r-1)[(r-1)(rk^3 - k^3 + 2) + k^2(r^2 - 3r + 1)] + k(5r^2 - 2r^3 - 3r + 1)]}$$

$$v = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(r-1)[(r-1)(rk^3 - k^3 + 2) + k^2(r^2 - 3r + 1)] + k(5r^2 - 2r^3 - 3r + 1)]}$$

$$\vdots \qquad\qquad \vdots$$

$$p = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(r-1)[(r-1)(rk^3 - k^3 + 2) + k^2(r^2 - 3r + 1)] + k(5r^2 - 2r^3 - 3r + 1)]} \quad (22)$$

Case 4: Same treatments missing in different replicates, but different blocks

For this case, the intra-block sum of squares error (SS_E) is;

$$SS_E = u^2 + v^2 + \left(\frac{(G'+u+v)^2}{B}\right) - \left(\frac{(R'_j+u)^2 + (R'_j+v)^2}{A}\right) - \left(\frac{(\tau'_i+u+v)^2}{r}\right) - \left(\frac{(C'_{jl}+u)^2 + (C'_{jl}+v)^2}{C}\right) + \left(\frac{(R'_c+u)^2 + (R'_c+v)^2}{D}\right) + \left(\frac{(S'_l+u)^2 + (S'_l+v)^2}{E}\right) + Q_4 \quad (23)$$

Differentiating (23) with respect to the missing plots, u and v, and equating to zero we have;

$$2u + \frac{2(G'+u+v)}{B} - \frac{2(R'_j+u)}{A} - \frac{2(\tau'_i+u+v)}{r} - \frac{2(C'_{jl}+u)}{C} + \frac{2(R'_c+u)}{D} + \frac{2(S'_l+u)}{E} = 0 \quad (24)$$

$$2v + \frac{2(G'+u+v)}{B} - \frac{2(R'_j+v)}{A} - \frac{2(\tau'_i+u+v)}{r} - \frac{2(C'_{jl}+v)}{C} + \frac{2(R'_c+v)}{D} + \frac{2(S'_l+v)}{E} = 0 \quad (25)$$

Solving simultaneously (24) and (25) to obtain estimates of the missing data, we have,

$$u = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[k(2k+1) + r(k^2+k-1)(r-3) - 2]} \quad (26)$$

$$v = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[k(2k+1) + r(k^2+k-1)(r-3) - 2]} \quad (27)$$

where H and I are as defined in Case 1.

Equations (26) and (27) give the non-iterative least square estimates for estimating two missing observations SRL and TRL designs when two different treatments are missing in different replicates, but different block.

Generally, for p missing values given case 4, we have,

$$u = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[k(2k+1) + r(k^2+k-1)(r-3) - 2]}$$

$$v = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[k(2k+1) + r(k^2+k-1)(r-3) - 2]}$$

$$\vdots \qquad\qquad \vdots$$

$$p = \frac{HR'_j + I\tau'_i - (r-1)(rk-k-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{(rk-k-1)[k(2k+1) + r(k^2+k-1)(r-3) - 2]} \quad (28)$$

3.4 Standard Errors of the Differences between Treatment Means

In rectangular lattice design without repetition, there are two simple formulae for estimating the standard error depending upon whether two treatments appear together in any one incomplete block or do not appear together.

The estimated standard error of the difference between the means of two treatments occurring together in an incomplete block is given by (29) as:

$$SE_1 = \sqrt{\varepsilon_e\left(\frac{k+1}{k}\right)\left\{\frac{2}{r} + \frac{k(k+1)}{r((r-1)(k^2-1)-k)}\right\}} \quad (29)$$

The estimated standard error of the difference between the means of two treatments not occurring together in an incomplete block is given by (30) as:

$$SE_2 = \sqrt{\varepsilon_e\left(\frac{k+2}{k}\right)\left\{\frac{2}{r} + \frac{k^3}{r(k+2)((r-1)(k^2-1)-k)}\right\}} \quad (30)$$

3.5 Estimation of Bias in Rectangular Lattice Designs without Repetitions

Statistical bias is a feature of a statistical technique or of its results whereby the expected value of the results differs from the true underlying quantitative parameter being estimated. Kabe and Gupta (2007) defined bias in estimated treatment sum of squares (SS) due to hypothesis as: Bias = [Estimated conditional error SS with estimated missing value] – [Minimum value of conditional error SS].

Consider simple and triple rectangular lattice designs with $k(k+1)$ treatments, r-replications and $k+1$ number of blocks where k is the block size. Suppose the yield $Y_{ijk} = u$ corresponding to the j^{th} replicate of i^{th} treatment in the l^{th} block is missing. Then u is estimated by (31) which is the expectation of u as

$$u = \hat{\mu} + \hat{\tau}_i + \hat{\theta}_j + \hat{\rho}_{l(j)} \tag{31}$$

where;

$$\hat{\mu} = \frac{(G' + u)}{B}\ ;\ \hat{\tau}_i = \frac{(\tau_i + u)}{r} - \frac{(G' + u)}{B}\ ;\ \hat{\theta}_j = \frac{(R'_j + u)}{A} - \frac{(G' + u)}{B}\ ;$$
$$\hat{\rho}_{l(j)} = \frac{(C'_{jl} + u)}{C} - \frac{(R'_c + u)}{D} - \frac{(S'_l + u)}{E}$$

and A, B, C, D, E and F retained their meaning as defined in (3.3)

By substituting the parameters of the model into equation (31), we have;

$$u = \frac{(G' + u)}{B} + \frac{(\tau'_i + u)}{r} - \frac{(G' + u)}{B} + \frac{(R'_j + u)}{A} - \frac{(G' + u)}{B} + \frac{(C'_{jl} + u)}{C} - \frac{(R'_c + u)}{D} - \frac{(S'_l + u)}{E} \tag{32}$$

Solving for u, we obtain the estimate as;

$$u = \frac{F(r-1)[rR'_j + A\tau'_i - G'] + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{F[(r-1)^2(k^2 + k - 1) - k]} \tag{33}$$

We shall now find the bias in the estimated treatment SS. Under **(t = 0),** the conditional error SS is given by Total SS – Block SS as

$$\left(\sum Y_{ijk}^2 + u^2 - \frac{(G' + u)^2}{B}\right) - \left(\frac{(C'_{jl} + u)^2}{C} - \frac{(R'_c + u)^2}{D} - \frac{(S'_l + u)^2}{E}\right) \tag{34}$$

Hence the estimated error SS denoted by C_E is;

$$C_E = \left(\sum Y_{ijk}^2 + \hat{u}^2 - \frac{(G' + \hat{u})^2}{B}\right) - \left(\frac{(C'_{jl} + \hat{u})^2}{C} - \frac{(R'_c + \hat{u})^2}{D} - \frac{(S'_l + \hat{u})^2}{E}\right) \tag{35}$$

where $\hat{u} = u$ in (33).

Equating to zero the derivative of the conditional error SS in (35) with respect to $\hat{u}$ and solving for $\hat{u}$. Denoting the resultant equation by u_o, we have

$$\frac{\partial C_E}{\partial \hat{u}} = 2\hat{u} - \frac{2(G' + \hat{u})}{B} - \frac{2(C'_{jl} + \hat{u})}{C} - \frac{2(R'_c + \hat{u})}{D} - \frac{2(S'_l + \hat{u})}{E} = 0 \tag{36}$$

$$u_o = \frac{F(r-1)G' + k[(r-1)(k+1)C'_{jl} - (r-1)R'_c - S'_l]}{F[(r-1)^2(rk^2 + rk - 1) - k]} \tag{37}$$

The minimum value of conditional error *SS* denoted by C_M is;

$$C_M = \left(\sum Y_{ijk}^2 + u_o^2 - \frac{(G' + u_o)^2}{B} \right) - \left(\frac{(C'_{jl} + u_o)^2}{C} - \frac{(R'_c + u_o)^2}{D} - \frac{(S'_l + u_o)^2}{E} \right) \quad (38)$$

The Bias in the estimated treatment SS is : *Bias* = C_E - C_M

$$Bias = \left\{ \left(\sum Y_{ijk}^2 + \hat{u}^2 - \frac{(G' + \hat{u})^2}{B} \right) - \left(\frac{(C'_{jl} + \hat{u})^2}{C} - \frac{(R'_c + \hat{u})^2}{D} - \frac{(S'_l + \hat{u})^2}{E} \right) \right\}$$
$$- \left\{ \left(\sum Y_{ijk}^2 + u_o^2 - \frac{(G' + u_o)^2}{B} \right) - \left(\frac{(C'_{jl} + u_o)^2}{C} - \frac{(R'_c + u_o)^2}{D} - \frac{(S'_l + u_o)^2}{E} \right) \right\} \quad (39)$$

Simplifying (39), we have

$$Bias = m \left[\frac{m'F[(r-1)(rk^2 + rk - 1) - k] - u^{\otimes}}{rk(r-1)(k+1)(rk-k-1)} \right] \quad (40)$$

where

$$u^{\otimes} = F(r-1)2G' + k[(r-1)(k+1)2C'_{jl} - (r-1)2R'_c - 2S'_l]\,;\; m = \hat{u} - u_o\,;\; m' = \hat{u} + u_o\,.$$

4. NUMERICAL EXAMPLE

In order to numerically illustrate the consistency and effectiveness of the proposed non-iterative least square formulae, a case study from Ford (1985) was adopted. The data presented in Table 3, show a 4×3 rectangular lattice design which was on the role of progesterone in stimulating sexual receptivity in estrogen-treated-ovariectomized gilts was used. Progesterone was administered simultaneously with estrogen. In this study, twelve gilts were treated with an optimal dosage of estradiol benzoate (EB). Progesterone treatment (600 micrograms/kg BW-1 X injection-1) on alternate days for a total of four injections produced serum concentrations of progesterone that were maximal at 9.4 *ng*/*ml* and remained greater than 1 *ng*/*ml* for fifteen days. Estradiol benzoate was administered twenty-two days after the first of these progesterone injections. When progesterone was administered concurrently with EB, the dosage was 100 micrograms/kg BW and produced a maximal serum progesterone concentration of 1.8 ng/ml four hours after treatment. Gilts were placed in randomized four-blocks evaluation pen of size three with a boar for five minutes on day 3 (*rep X*), day 4 (*rep Y*) and day 5 (*rep Z*) after EB treatment. Trait of interest was the total number of mounts by the boar.

Table 3
Rectangular Lattice Data on Estrogen-Treated Gilts
(Treatment Numbers are Enclosed in Parentheses)

Blocks	**Mounts, Number/5min**		
	Rep X		
4	(10)7	(12)5	(11)4
1	(2)5	(3)2	(1)1
3	(7)4	(9)3	(8)4
2	(4)3	(5)3	(6)2
	Rep Y		
4	(3)6	(6)6	(9)6
2	(1)4	(11)2	(8)3
3	(12)3	(2)3	(5)1
1	(10)1	(4)4	(7)3
	Rep Z		
1	(8)7	(6)5	(12)4
2	(9)4	(10)4	(2)3
3	(11)1	(3)2	(4)2
4	(5)5	(1)3	(7)4

5. ANALYSIS AND RESULTS

The methodology steps on how the proposed formulae were used are thus; suppose that some observation(s) for certain treatment(s) in the block(s) of any replication is (are) missing at random, given SRL or TRL design without repetition respectively; computationally obtain the corresponding quantities for estimation; substitute the quantities into each proposed formula, depending on whether it is a SRL or TRL design to estimate the missing value(s); obtain the estimated standard errors and p-value(s) after adjustment has been made for the bias using (40).

Suppose in Table 3, the single observation 7 for treatment (10) in block 4 of rep. X had been missing given a single case, it is estimated using (4). For two values missing in group X and Y given case 1, 2, 3 and 4 respectively, we suppose that the observation 7 for treatment (10) in block 4 of rep. X and the observation 6 for treatment (3) in block 4 of rep. Y had been missing given case 1; the observation 5 for treatment (2) in block 1 of rep. X and the observation 2 for treatment (3) in block 1 of rep. X had been missing given case 2; the observation 4 for treatment (10) in block 2 of rep. Z and the observation 5 for treatment (5) in block 4 of rep. Z had been missing given case 3; and the observation 4 for treatment (1) in block 2 of rep. Y and the observation 3 for treatment (1) in block 4 of rep. Z had been missing given case 4, the non-iterative least square estimates are obtained using (8) and (9) for case 1, (14) and (15) for case 2, (20) and (21) for case 3, and (26) and (27) for case 4 while the standard errors are obtained using (29) and (30). The computations and summary of the results are presented in (5.1), (5.2) and Table 4 respectively.

5.1 Computational Procedures for SRL Design using the Proposed Formulae

Estimates for single missing observation

$$u = \frac{1}{16}[2(101.4) + 3(-39)] = 5.36\,;\ SE_1 = \sqrt{(2.510)(1.614)} = 2.01274\,;$$

$$SE_2 = \sqrt{(2.7381)(1.7597)} = 2.15459\,;\ p - value = 0.0001$$

Estimates for two missing observation

Case 1: Different treatments missing in different replicates, but same block

$$u = \frac{1}{21}[2(113.55) + 3(-45)] = 4.38571\ ;\ v = \frac{1}{21}[2(129.1) + 3(-37)] = 7.00952\,;$$

$$SE_1 = \sqrt{(3.6210)(3.4389)} = 3.52878\,;\ SE_2 = \sqrt{(3.9101)(3.700)} = 3.8036\,;$$
$$p - value = 0.0001$$

Case 2: Different treatments missing in same replicates, but same block

$$u = \frac{1}{12}[2(93.08) + 3(-38)] = 6.01333\,;\ v = \frac{1}{12}[2(78.77) + 3(-41)] = 2.87833\,;$$

$$SE_1 = \sqrt{(3.4341)(3.437)} = 3.43555\,;\ SE_2 = \sqrt{(3.731)(3.429)} = 3.57681\,;$$
$$p - value = 0.0001$$

Case 3: Different treatments missing in same replicates, but different blocks

$$u = \frac{1}{17}[2(121.1) + 3(-38)] = 7.54118\,;\ v = \frac{1}{17}[2(85.23) + 3(-38)] = 3.3212\,;$$

$$SE_1 = \sqrt{(4.5364)(3.659)} = 4.07415\,;\ SE_2 = \sqrt{(4.5367)(3.9324)} = 4.22376\,;$$
$$p - value = 0.0001$$

Case 4: Same treatments missing in different replicates, but different blocks

$$u = \frac{1}{6}[2(79.52) + 3(-40)] = 6.50667\,;\ v = \frac{1}{6}[2(70.41) + 3(-37)] = 4.97\,;$$

$$SE_1 = \sqrt{(4.911)(3.1854)} = 3.95519\,;\ SE_2 = \sqrt{(4.920)(3.2434)} = 3.99469\,;$$
$$p - value = 0.0001$$

Bias for SRL design without repetitions

$$Bias = \left[(-1.55) \times \frac{(25.05 \times 35) - 903.6}{48} \right] = 0.867031$$

5.2 Computational Procedures for TRL Design using the Proposed Formulae

Estimates for single missing observation

$$u = \frac{1}{205}[10(186.3) + 3(-58)] = 8.23902 \;\; SE_1 = \sqrt{(1.301)(0.3121)} = 0.637214;$$

$$SE_2 = \sqrt{(1.399)(0.3525)} = 0.702245; \;\; p-value = 0.0001$$

Estimates for two missing observation

Case 1: Different treatments missing in different replicates, but same block

$$u = \frac{1}{218}[10(223.4) + 3(-51)] = 9.54587; \;\; v = \frac{1}{218}[10(140.7) + 3(-47)] = 5.80734;$$

$$SE_1 = \sqrt{(2.258)(0.343)} = 0.880053; \;\; SE_2 = \sqrt{(2.355)(0.3434)} = 0.899281;$$

$$p-value = 0.0001$$

Case 2: Different treatments missing in same replicates, but same block

$$u = \frac{1}{170}[10(111.23) + 3(-41)] = 5.81941; \;\; v = \frac{1}{170}[10(63.7) + 3(-36)] = 3.11176;$$

$$SE_1 = \sqrt{(2.411)(0.2623)} = 0.795239; \;\; SE_2 = \sqrt{(2.9394)(0.2484)} = 0.854486;$$

$$p-value = 0.0001$$

Case 3: Different treatments missing in same replicates, but different blocks

$$u = \frac{1}{191}[10(114.7) + 3(-43)] = 5.32984; \;\; v = \frac{1}{191}[10(144.8) + 3(-44)] = 6.89005;$$

$$SE_1 = \sqrt{(1.940)(0.430)} = 0.913345; \;\; SE_2 = \sqrt{(1.551)(0.60)} = 0.964676;$$

$$p-value = 0.0001$$

Case 4: Same treatments missing in different replicates, but different blocks

$$u = \frac{1}{95}[10(59.5) + 3(-44)] = 4.87368; \;\; v = \frac{1}{95}[10(44.4) + 3(-42)] = 3.34737;$$

$$SE_1 = \sqrt{(1.5220)(0.375)} = 0.75548; \;\; SE_2 = \sqrt{(1.5251)(0.4001)} = 0.781148;$$

$$p-value = 0.0001$$

Bias for TRL design without repetitions

$$Bias = \left[(8.99) \times \frac{(35.4 \times 261.76) - 9205}{360}\right] = 1.5309$$

Table 4
Summary of Results for Estimating Missing Values in SRL and TRL Designs without repetition

	Estimates		One Obs.	Two Observations			
				Case 1	**Case 2**	**Case 3**	**Case 4**
SRL design	Complete data	Value	7	7 6	5 2	4 5	4 3
		SE_1	2.3751	2.3751	2.3751	2.3751	2.3751
		SE_2	2.3886	2.3886	2.3886	2.3886	2.3886
		P-value	0.0001	0.0001	0.0001	0.0001	0.0001
	Estimated	Value	5.36	4.39 7.01	6.01 2.88	7.54 3.32	6.51 4.97
		SE_1	2.0132	3.5289	3.4361	4.0742	3.9552
		SE_2	2.1554	3.8031	3.5770	4.2238	3.9947
		P-value	0.0001	0.0001	0.0001	0.0001	0.0001
TRL design	Complete data	Value	7	7 6	5 2	4 5	4 3
		SE_1	1.8323	1.8323	1.8323	1.8323	1.8323
		SE_2	1.9225	1.9225	1.9225	1.9225	1.9225
		P-value	0.0001	0.0001	0.0001	0.0001	0.0001
	Estimated	Value	8.24	9.55 5.81	5.82 3.11	5.33 6.89	4.87 3.35
		SE_1	0.6371	0.8801	0.7952	0.9133	0.7556
		SE_2	0.7022	0.8993	0.8545	0.9650	0.7811
		P-value	0.0001	0.0001	0.0001	0.0001	0.0001

Table 4 showed the results from the analysis of the complete data using the proposed non-iterative formulae for when one or two observations are missing. The estimated standard errors and p-values were obtained after adjusting for the bias effect using the computational value of 0.8670 and 1.5309 for SRL and TRL designs respectively.

6. DISCUSSION

The empirical results from the analysis of the proposed formulae for estimating missing value(s) in SRL and TRL designs without repetition are presented in Table 4. For the estimation of one and two missing value(s) in SRL and TRL designs respectively, we minimized the intra-block error sum of squares with respect to the missing plot(s), and solve for the missing plot(s) to obtain the estimate(s) for each basic design.

The estimated values, standard errors and p-values have been evaluated to show the application and suitability of the proposed formulae for estimating missing data in SRL and TRL designs. Results of the empirical analysis of the proposed formulae and the complete data for SRL and TRL designs respectively showed that the proposed formulae are appropriate and suitable for the estimation of missing value(s) in SRL and TRL designs without repetition since the estimated value(s) are significantly approximate to the complete data value(s) as presented in Table 4.

7. CONCLUSION

In this paper, a non-iterative least square technique was used to propose formulae for the estimation of missing data in SRL and TRL designs without repetition respectively when one or more observations are missing using the intra-block information. The estimated values, standard errors and p-values were evaluated to compare the empirical results of the proposed formulae with those of the complete data, and to show the applicability of the proposed formulae. Results obtained showed that the estimated values of the proposed formulae were significantly approximate to the complete data values which indicate the suitability and applicability of the proposed formulae in the estimation of missing data in SRL and TRL designs respectively. It is therefore recommended that whenever missing data occur in any designed experiment involving the use of SRL or TRL design without repetitions respectively, the proposed formulae should be adopted in the estimation of the missing value(s).

REFERENCES

1. Ahmed, L.A. (2016). Missing values estimation comparisons in split-plot design. *International Journal of Computer and Information Technology*, 5(3), 337-344.
2. Allison, P.D. (2001). *Missing data*. London: Sage Publications, Inc.
3. Alrweili, H. Georgiou, S. and Stylianou, S. (2019). Robustness of response surface designs to missing data. *Quality and Reliability Engineering International*, 35, 1288-1296.
4. Azadeh, A., Asadzadeh, S.M., Jafari-Marandi, R., Nasari-Shirkouhi, S., Baharian, K.G. and Talebi, S. (2013). Optimum estimation of missing values in randomized complete block design by genetic algorithm. *Knowledge Based System*, 37, 37-47.
5. Bhatra, C.N. Ch. and Dharamyadav, T. (2013). Estimation of missing observation in randomized block design. *International Journal of Technology and Engineering Science*, 1(6), 618-621.
6. Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs.* London: Wiley.
7. Coertjens, L., Donche, V., De Maeyer S., Vanthournout, G. and Van, P. (2017). To what degree does the missing-data technique influence the estimated growth in learning strategies overtime? A Tutorial Example of the Sensitivity Analysis for Longitudinal Data. *PloS One*, 12(9), 1-21.
8. Ford, J.J. (1985). Reevaluation of the role of progesterone in stimulating sexual receptivity in estrogen-treated gilts. *Journal of Animal Science*, 61(1), 36-43.
9. Gad, A.M. and Ahmed, A.S. (2006). Analysis of longitudinal data with intermittent missing values using the stochastic EM algorithm. *Computational Statistics and Data Analysis*, 50(10), 2702-2714.
10. Geeta, C., Vasudha, V. and Jayanthi, R. (2017). A comparison of multiple imputation methods for data with missing values. *Indian Journal of Science and Technology*, 10(19), 1-7.
11. Harshbarger, B. (1946). Preliminary report on rectangular lattices. *Biometrics Bulletin*. 2, 115-119.
12. Harshbarger, B. (1947). Rectangular lattices. Virginia Agricultural Experiment Station, *Memoir* I.

13. Harshbarger, B. (1949). Triple rectangular lattices. *Biometrics*, 5, 1-13.
14. Jakobson, J.C., Gluud, C., Wetterslev, J. and Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomized clinical trials- a practical guide with flowcharts. *BMC Medical Research Methodology Series*, 17(1), 1-10.
15. Kabe, D.G. and Gupta, A.K. (2007). *Experimental Designs: Exercises and Solutions*. Springer-Verlag New York.
16. Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.
17. Meiller, Y., Guillemont, J., Beishuizen C., Richard, E., Kivipelto, M. and Andrieu, S. (2017). On missing data encodings for information systems. Conference proceeding: *Americas Conference on Information System*, (AMCIS 2016) San Diego, USA.
18. Nuryazmin, A.Z., Abdul, A.J. and Nora, M. (2015). A comparison of various imputation methods for missing values in air quality data. *Sains Malaysiana*, 44(3), 449-456.
19. Okereke, E.W., Ekpenyong, E.J. and Nwaogu, C. (2018). Missing value estimation in a nested-factorial design with three factors. *Trends Journal of Sciences Research*, 3(1), 10-17.
20. Ossai, E.O. and Oladugba, A.V. (2018). The estimation of missing values in rectangular lattice designs. *Journal of Modern Applied Statistical Methods*, 17(2), 1-23.
21. Schmitt, P., Mandel, J. and Guedj, M. (2015). A Comparison of six methods of missing data imputation. *Journal of Biometrics and Biostatistics*, 6(1), 1-6.
22. Singh, Y. (2019a). On the analysis of simple rectangular lattice design. https://www.researchgate.net/publication/332319733_On_the_Analysis_of_Simple_Rectangular_Lattice_Design
23. Singh, Y. (2019a). On the analysis of triple rectangular lattice design. https://www.researchgate.net/publication/332319808_A_Note_on_Triple_Rectangular_Lattice_Design
24. Subramani, J. (1991a). Non-iterative least squares estimation of missing values in Graeco-Latin square designs. *Biometrical Journal*, 33(6), 763-769.
25. Subramani, J. (1991b). Non-iterative least squares estimation of missing values in replicated Latin square designs. *Biometrical Journal*, 33(8), 999-1011.
26. Subramani, J. (1994). Non-iterative least squares estimation of missing values in cross-over designs without residual effect. *Biometrical Journal,* 36(3), 285-292.
27. Sullivan, T.R., Yelland L.N., Leek, K.J., Ryan, P. and Satter A.B. (2017). Treatment of missing data in follow-up studies of randomized controlled trials: *A Systematic Review of the Literature,* 14(4), 387-395.
28. Takahashi, M. (2017). Statistical inference in missing data by MCMC and non-MCMC multiple imputation algorithm: assessing the effect of between-imputation iterations. *Data Science Journal*, 16(37), 1-17.
29. Yize, Z. and Qi, L. (2017). Variable selection in the presence of missing data: imputation-based methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(5), e1402.