Trends Journal of Sciences Research

Trends Journal of Sciences Research, 2018, Volume 3, Issue 1, 10-17
http://www.tjsr.org/journal/index.php/tjsr
DOI: 10.31586/Statistics.0301.02

Missing Value Estimation in a Nested-Factorial Design with Three Factors

Emmanuel W. Okereke*, **Emmanuel J. Ekpenyong**, **Chukwuma Nwaogu**

Department of Statistics, Michael Okpara University, Umudike, Nigeria
*Corresponding author: emmastat5000@yahoo.co.uk

OPEN ACCESS Research Article

How to cite this article: Okereke, E. W., Ekpenyong, E. J., & Nwaogu, C. (2018). Missing Value Estimation in a Nested-Factorial Design with Three Factors. *Trends Journal of Sciences Research*, 3(1), 10-17 https://doi.org/10.31586/Statistics.0301.02

Received: June 01, 2018

Accepted: July 24, 2018

Published: July 25, 2018

Abstract When faced with unbalanced data, it is often necessary to estimate the necessary missing values before the application of the analysis of variance technique. Previous studies have shown that diferent designs require diferent missing value estimators. With the introduction of some relatively new statistical designs, it has become expedient to derive missing value estimators for such designs. In this study, least squares estimators of missing values in a three-factor nested-factorial design are derived. Properties of the estimators are equally determined. A numerical example is given to show the application of the theoretical results obtained in this paper. Our empirical results establish the appropriateness of the missing value estimation method presented in this study.

Keywords: *Nested-Factorial Design, Non-Iterative Least Squares Estimation, Bias, Analysis of Vari- Ance, Missing Value*

Introduction

Comparative experiments are often inevitable in many scientific studies. They serve as the means of generating data. Therefore, care is usually taken to ensure that such experiments are properly conducted. Before carrying out a comparative experiment, an experimenter may have to adopt a suitable experimental(statistical) design. Several statistical designs have been proposed for use under certain experimental conditions[1,2].

Data collected in the course of a well design experiment need to be analysed in order to provide answers to research questions under consideration. If quantitative data are classified according to three or more treatments or levels of at least two factors, an analysis of variance (ANOVA) tech-nique may be applied. Different statistical designs require different analysis of variance techniques. For instance, one-way ANOVA is applicable to data collected using the completely randomised design.

No matter how carefully planned and conducted an experiment is, there might be a case of unbalanced data. ANOVA models were originally developed for balanced data. The problem of performing analysis of variance on unbalanced data can be handled by first estimating the missing values and using the estimates in place of the missing observations. The resulting data, comprising the actual observations and the estimates of the missing values are then analysed. Following the novel works of [3,4], least squares estimators of missing values in a number of statistical designs, namely, Randomised Block Design[5], General Incomplete Block

Design [6], Latin Square Design [7,8], Graeco-Latin Square Design [9], F-Square Design [10], Cross-Over Design [11] and Split-Plot Design [12] have been derived. The purpose of this paper is to derive the least squares estimators of missing values in a nested factorial design. Statistical properties of the estimators are equally investigated.

Review of two-stage nested design and three-factor nested factorial design

Nested designs among other statistical designs are frequently used in agricultural, ecological, medical and industrial experimental processes [13]. There are generally classified in accordance with the number of factors used in the experiment. For instance, in a experimental situation where two factors (say A and B) are being considered such that each level of B is combined with only one level of A, we say B is nested in A. The resulting design is called a two-stage nested design. The linear statistical model for a balanced two-stage nested design, may be written as

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}, i = 1, 2, \cdots, a, j = 1, 2, \cdots, b, k = 1, 2, \cdots, r \quad (1)$$

where Y_{ijk} is the kth observation at the jth level of B nested in the ith level of A, μ is the grand mean, α_i is the effect of ith level of factor A, $\beta_{j(i)}$ is the effect of jth level of factor B nested within ith level of factor A and $\epsilon_{k(ij)}$ is the random error term such that $\epsilon_{k(ij)} \sim N(0, \sigma_e^2)$. The nature of this design makes it impossible for one to examine the main effect of factor B and the interaction between the two factors [14]. In a two-stage nested design, the hypotheses to be tested, depend on whether the two factors are fixed or random or we have a combination of fixed and random factors. In these three cases, the partitioning of the total variation into recognised sources of variation remains the same. Let SS_A, $SS_{B(A)}$ and SS_{E_1} denote the sum of squares due to factor A, sum of squares due to factor B within the levels of factor A and sum of squares due to error respectively. The total sum of squares (SS_T) is partitioned as follows:

$$SS_T = SS_A + SS_{B(A)} + SS_{E_1} \quad (2)$$

where $SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} X_{ijk}^2 - \frac{1}{abr} X_{...}^2$, $SS_A = \frac{1}{br} \sum_{i=1}^{a} X_{i..}^2 - \frac{1}{abr} X_{...}^2$,

$SS_{B(A)} = \frac{1}{r} \sum_{i=1}^{a} \sum_{j=1}^{b} X_{ij.}^2 - \frac{1}{br} X_{...}^2$, $SS_{E_1} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} X_{ijk}^2 - \frac{1}{r} \sum_{i=1}^{a} \sum_{j=1}^{b} X_{ij.}^2$
and $X_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} X_{ijk}$. Table 1 comprises the various cases of two-stage nested design.

Table 1. Test Statistics and Rejection Criteria for Various Cases of Two-Stage Nested Designs

Factor	Type of effects	Hypotheses	F-Ratio	Reject H_0 if
A	fixed	$H_0 : \alpha_i = 0$ vs $H_1 : \alpha_i \neq 0$	$F_1 = \frac{MS_A}{MS_E}$	$F_1 > F_{\alpha,(a-1),ab(n-1)}$
B	fixed	$H_0 : \beta_{j(i)} = 0$ vs $H_1 : \beta_{j(i)} \neq 0$	$F_2 = \frac{MS_{B(A)}}{MS_E}$	$F_2 > F_{\alpha,a(b-1),ab(n-1)}$
A	random	$H_0 : \sigma_\alpha^2 = 0$ vs $H_1 : \sigma_\alpha^2 \neq 0$	$F_3 = \frac{MS_A}{MS_{B(A)}}$	$F_3 > F_{\alpha,(a-1),a(b-1)}$
B	random	$H_0 : \sigma_{\beta_{j(i)}}^2 = 0$ vs $H_1 : \sigma_{\beta_{j(i)}}^2 \neq 0$	$F_2 = \frac{MS_{B(A)}}{MS_E}$	$F_2 > F_{\alpha,a(b-1),ab(n-1)}$
A	fixed	$H_0 : \alpha_i = 0$ vs $H_1 : \alpha_i \neq 0$	$F_3 = \frac{MS_A}{MS_{B(A)}}$	$F_3 > F_{\alpha,(a-1),a(b-1)}$
B	random	$H_0 : \sigma_{\beta_{j(i)}}^2 = 0$ vs $H_1 : \sigma_{\beta_{j(i)}}^2 \neq 0$	$F_2 = \frac{MS_{B(A)}}{MS_E}$	$F_2 > F_{\alpha,a(b-1),ab(n-1)}$
A	random	$H_0 : \sigma_\alpha^2 = 0$ vs $H_1 : \sigma_\alpha^2 \neq 0$	$F_1 = \frac{MS_A}{MS_E}$	$F_1 > F_{\alpha,(a-1),ab(n-1)}$
B	fixed	$H_0 : \beta_{j(i)} = 0$ vs $H_1 : \beta_{j(i)} \neq 0$	$F_2 = \frac{MS_{B(A)}}{MS_E}$	$F_2 > F_{\alpha,a(b-1),ab(n-1)}$

In Table 1, α is the level of significance, $MS_A = \frac{SS_A}{a-1}$, $MS_{B(A)} = \frac{SS_{B(A)}}{a(b-1)}$ and $MS_E = \frac{SS_{E_1}}{ab(n-1)}$.

A nested-factorial design is a statistical design that involves both crossed and nested factors. Suppose that in a three-factor nested-factorial design, factors A, B and C have a levels, b levels and c levels respectively. If the b levels of factor B are nested within a levels of factor A and c levels of factor C are crossed with a levels of factor A and b levels of factor B, we may consider the linear model:

$$Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk(i)} + e_{l(ijk)} \tag{3}$$

In (3), μ is the grand mean, α_i is the effect of the ith level of factor A, $\beta_{j(i)}$ is the effect of jth level of factor B nested within ith level of factor A, γ_k is the effect attributable to kth level of factor C, $(\alpha\gamma)_{ik}$ is the effect of the interaction of ith level of factor A and kth level of factor C, $(\beta\gamma)_{jk(i)}$ represents the interaction effect of the kth level of factor C and jth level of factor B within the ith level of factor A and $e_{l(ijk)}$ is the error term.

The total sum of squares (SS_T) corresponding to (3), is partitioned as follows:

$$SS_T = SS_A + SS_{B(A)} + SS_C + SS_{AC} + SS_{BC(A)} + SS_E \tag{4}$$

where

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} X_{ijkl}^2 - \frac{X_{....}^2}{abcn} \tag{5}$$

$$SS_A = \frac{1}{bcn}\sum_{i=1}^{a} X_{i...}^2 - \frac{X_{....}^2}{abcn} \tag{6}$$

$$SS_{B(A)} = \frac{1}{cn}\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij..}^2 - \frac{1}{bcn}\sum_{i=1}^{a} X_{i...}^2 \tag{7}$$

$$SS_C = \frac{1}{abn}\sum_{k=1}^{c} X_{..k.}^2 - \frac{X_{....}^2}{abcn} \tag{8}$$

$$SS_{AC} = \frac{1}{bn}\sum_{i=1}^{a}\sum_{k=1}^{c} X_{i.k.}^2 - \frac{1}{bcn}\sum_{i=1}^{a} X_{i...}^2 - \frac{1}{abn}\sum_{k=1}^{c} X_{..k.}^2 + \frac{X_{....}^2}{abcn} \tag{9}$$

$$SS_{BC(A)} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} X_{ijk.}^2 - \frac{1}{cn}\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij..}^2 - \frac{1}{bn}\sum_{i=1}^{a}\sum_{k=1}^{c} X_{i.k.}^2 + \frac{1}{bcn}\sum_{i=1}^{a} X_{i...}^2 \tag{10}$$

$$X_{....} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} X_{ijkl} \tag{11}$$

and

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n} X_{ijkl}^2 - \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} X_{ijk.}^2 \tag{12}$$

Table 2. Test Statistics and Rejection Criteria based on the three-factor nested-factorial design

Factor	Type of effects	Hypotheses	F-Ratio	Reject H_0 if
A	fixed	$H_0: \alpha_i = 0$ vs $H_1: \alpha_i \neq 0$	F_4	$F_4 > F_{\alpha,(a-1),abc(n-1)}$
B(A)	fixed	$H_0: \beta_{j(i)} = 0$ vs $H_1: \beta_{j(i)} \neq 0$	F_5	$F_5 > F_{\alpha,a(b-1),abc(n-1)}$
C	fixed	$H_0: \gamma_k = 0$ vs $H_1: \gamma_k \neq 0$	F_6	$F_6 > F_{\alpha,c-1,abc(n-1)}$
A × C		$H_0: (\alpha\gamma)_{ik} = 0$ vs $H_1: (\alpha\gamma)_{ik} \neq 0$	F_7	$F_7 > F_{\alpha,(a-1)(c-1),abc(n-1)}$
B × C(A)		$H_0: (\beta\gamma)_{jk(i)} = 0$ vs $H_1: (\beta\gamma)_{jk(i)} \neq 0$	F_8	$F_8 > F_{\alpha,a(b-1)(c-1),abc(n-1)}$
A	fixed	$H_0: \alpha_i = 0$ vs $H_1: \alpha_i \neq 0$	F_9	$F_9 > F_{\alpha,(a-1),a(b-1)}$
B(A)	random	$H_0: \sigma^2_{\beta(\alpha)} = 0$ vs $H_1: \sigma^2_{\beta(\alpha)} \neq 0$	F_5	$F_5 > F_{\alpha,a(b-1),abc(n-1)}$
C	fixed	$H_0: \gamma_k = 0$ vs $H_1: \gamma_k \neq 0$	F_{10}	$F_{10} > F_{\alpha,c-1,a(b-1)(c-1)}$
A × C		$H_0: (\alpha\gamma)_{ik} = 0$ vs $H_1: (\alpha\gamma)_{ik} \neq 0$	F_{10}	$F_{11} > F_{\alpha,(a-1)(c-1),a(b-1)(c-1)}$
B × C(A)		$H_0: \sigma^2_{\beta\gamma(\alpha)} = 0$ vs $H_1: \sigma^2_{\beta\gamma(\alpha)} \neq 0$	F_{12}	$F_{12} > F_{\alpha,a(b-1)(c-1),abc(n-1)}$

where $F_4 = \frac{MS_A}{MS_E}$, $F_5 = \frac{MS_{B(A)}}{MS_E}$, $F_6 = \frac{MS_C}{MS_E}$, $F_7 = \frac{MS_{AC}}{MS_E}$, $F_8 = \frac{MS_{BC(A)}}{MS_E}$, $F_9 = \frac{MS_A}{MS_{BC(A)}}$, $F_{10} = \frac{MS_C}{MS_{BC(A)}}$ $F_{11} = \frac{MS_{AC}}{MS_{BC(A)}}$ and $F_{12} = \frac{MS_{BC(A)}}{MS_E}$.

Main Results

In this section, we derive least squares estimators of missing values in a three-factor nested-factorial design under several conditions. Theorem 1 provides the estimators of s missing values within the same cell in nested-factorial design.

Theorem 3.1. *Suppose there are n numbers of observations per each combination of a level of each of factors A, B, and C in a nested-factorial design. Assume s of the r observations are missing. Let the least squares estimators of the missing values be $M_1, M_2, M_3, \cdots, M_s$. The estimators are all equal to the arithmetic mean of the $(n-s)$ observations remaining in the cell that contains the missing values.*

Proof. From (11), we have

$$SS_E = \sum_{y=1}^{s} M_y^2 + R - \frac{(X'_{....} + \sum_{y=1}^{s} M_y)^2}{n}$$

where R is the sum of all the terms independent of $M_1, M_2, M_3, \cdots, M_s$. The partial derivatives of SS_E with respect to $M_1, M_2, M_3, \cdots, M_s$ satisfy the equations $\frac{\partial SS_E}{\partial M_y} = 2M_y - \frac{2(X'_{....} + \sum_{y=1}^{s} M_y)}{n}, 1, 2, 3, \cdots, s$. Equating to zero the partial derivative of SS_E with respect to each of $M_1, M_2, M_3, \cdots, M_s$ leads to the following system of linear equations:

$$\mathbf{C}_{s\times s}\mathbf{M}_{s\times 1} = \mathbf{X}_{s\times 1}$$
$$\mathbf{M}_{s\times 1} = \mathbf{C}_{s\times s}^{-1}\mathbf{X}_{s\times 1} \tag{13}$$

where $\mathbf{C}_{s\times s} = \begin{pmatrix} n-1 & -1 & -1 & \cdots & -1 \\ -1 & n-1 & -1 & \cdots & -1 \\ -1 & -1 & n-1 & \cdots & -1 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ -1 & -1 & -1 & \cdots & n-1 \end{pmatrix}$, $\mathbf{M}_{s\times 1} = \begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ . \\ . \\ . \\ M_s \end{pmatrix}$, $\mathbf{X}_{s\times 1} = \begin{pmatrix} X'_{ijk.} \\ X'_{ijk.} \\ X'_{ijk.} \\ . \\ . \\ . \\ X'_{ijk.} \end{pmatrix}$

and $X'_{ijk.}$ is the sum of the $(n-s)$ observations that are originally available in the cell. Next, we solve for $\mathbf{M}_{s\times 1}$ in (13) using the principle of mathematical induction. Before obtaining the general solution of (13), we shall solve (13) when s=1, 2 and 3. If $s = 1$, we have

$$M_1 = \frac{X'_{ijk.}}{n-1} \tag{14}$$

For $s = 2$,

$$\begin{pmatrix} n-1 & -1 \\ -1 & n-1 \end{pmatrix} \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} = \begin{pmatrix} X'_{ijk.} \\ X'_{ijk.} \end{pmatrix} \tag{15}$$

Solving (13) for M_1 and M_2 leads to

$$M_1 = M_2 = \frac{X'_{ijk.}}{n-2} \tag{16}$$

With $s = 3$, the following equation is satisfied:

$$\begin{pmatrix} M_1 \\ M_2 \\ M_3 \end{pmatrix} = \begin{pmatrix} n-1 & -1 & -1 \\ -1 & n-1 & -1 \\ -1 & -1 & n-1 \end{pmatrix}^{-1} \begin{pmatrix} X'_{ijk.} \\ X'_{ijk.} \\ X'_{ijk.} \end{pmatrix} \tag{17}$$

$$\Longrightarrow M_1 = M_2 = M_3 = \frac{X'_{ijk.}}{n-3} \tag{18}$$

Consequently, the solution of (13) is

$$M_y = \frac{X'_{ijk.}}{n-s}, y = 1, 2, 3, \cdots, s \qquad QED \tag{19}$$

It may happen that the missing values we wish to estimate belong to different cells.

Theorem 3.2. *Let $V_1, V_2, \cdots, V_q$ denote least squares estimators of missing observations in q different cells in a nested-factorial design with three factors, such that in each of the cells only one value is missing. Let the number of observations originally available in each of the q cells be $n-1$. Denote the totals of observations originally available in the cells by $X'_{i^{(e)}j^{(e)}k^{(e)}l^{(e)}}, e = 1, 2, 3, \cdots, q$.*

Then $V_e = \frac{X'_{i(e)j(e)k(e)l(e)}}{n-1}, e = 1, 2, 3, \cdots, q.$

Proof. Using (11), we obtain

$$SS_E = \sum_{e=1}^{q} V_e^2 + R' - \frac{\sum_{e=1}^{q}(V_e + X_{i(e)j(e)k(e)l(e)})^2}{n}$$

where R' is the sum of all the terms independent of $V_1, V_2, V_3, \cdots, V_q$. The partial derivatives of SS_E with respect to $V_1, V_2, V_3, \cdots, V_q$ satisfy the equations $\frac{\partial SS_E}{\partial V_e} = 2V_e - \frac{2(V_e + X_{i(e)j(e)k(e)l(e)})}{n}$, $V_1, V_2, v_3, \cdots, V_q$. On equating $\frac{\partial SS_E}{\partial V_e}$ to zero and solving the resulting equation, we have

$$V_e = \frac{X'_{i(e)j(e)k(e)l(e)}}{n-1}, e = 1, 2, 3, \cdots, q \tag{20}$$

This completes the proof.

Other cases of missing values in a nested-factorial design with three factors may be frequently encountered. For instance, two or more of the q missing values may belong to the same cell. The fact remains that least squares etimators of such missing values can be easily derived using similar procedures to those in Theorem 3.1 and 3.2.
It has been argued by many authors that when a missing value is estimated , as it is the case in this study, the treatment sum of squares is biased. The bias in sum of squares due to factor C, which may be encountered when a missing value in the design under consideration, is estimated using (13), is given in Theorem 3.

Theorem 3.3. *Let a missing value in a three-factor nested-factorial design be estimated using (14). If when we ignore the classification of the observations based on factor C, we obtain a two-way nested design in which the number of observations originally available in the cell containing the missing observation is* $r-1$*. Let the estimator of the missing observation in the resulting nested design be* U*. In testing the hypothesis* $H_0 : \gamma_k = 0$*, the sum of squares due to factor C is positively biased. The bias is* $B = \left(\frac{r-1}{r}\right)(M_1 - U)^2$*.*

Proof. By substituting M_1 and U into SS_{E_1}, the bias is obtained as follows:

$$\begin{aligned} B &= M_1^2 - \frac{(X'_{ij.} + M_1)^2}{r} - U^2 - \frac{(X'_{ij.} + U)^2}{r} \\ &= \left(\frac{r-1}{r}\right)(M_1 - U)^2 \end{aligned} \tag{21}$$

In the case of one missing value in a three-factor nested factorial design, the missing value is estimated using (14) and adjustment for bias in SS_C is made by subtracting B from SS_C [15] In general, if there are two or more missing values, the estimates of the values are found using the appropriate formulae based on the nested and nested-factorial designs. These estimates are then used in place of the corresponding missing values and the analysis of variance for both nested and nested-factorial designs are conducted. The corresponding SS_E and SS_{E_1} are computed. As a consequence, the corrected sum of squares due to factor C is [16]

$$\text{Corrected } SS_C = SS_{E_1} - SS_E \tag{22}$$

Numerical Example

Numerical illustrations made in this section are based on the assembly time data from [2]. The data were collected in an experiment in which three-factor nested factorial design was applied. Of interest in the experiment are the three factors operators, layouts and fixtures, which have four levels, two levels and three levels respectively . Among the three factors considered in the experiment, operators are nested under levels of layouts. It shall be noted that the four operators selected for Layout 1 are different from the four operators selected for Layout 2. Moreover, the operators are randomly selected, justifying the use of the mixed effects analysis of variance model. As shown in Table 3, the third factor fixtures and layouts are subjected to a factorial arrangement.

Table 3. The Assembly Time Data

Operators	Layout 1				Layout 2			
	1	2	3	4	1	2	3	4
Fixture 1	22	23	28	25	26	27	28	24
	24	24	29	23	28	25	25	23
Fixture 2	30	29	30	27	29	30	24	28
	27	28	32	25	28	27	23	30
Fixture 3	25	24	27	26	27	26	24	28
	21	22	25	23	25	24	27	27

Source: Montgomery(2013)

For easy reference to each observation in Table 3, the observations will be expressed in X_{ijkl} notation, where $i = 1, 2, 3, 4, j = 1, 2, k = 1, 2, 3, 4, l = 1, 2$. For instance, $X_{1111} = 22$ refers to the first observation in the cell corresponding to Operator 1, Layout 1 and Fixture 1. The data in Table 1 have been analysed in [2]. However, for reference purposes, we consider the ANOVA results in Table 4.

Table 4. Analysis of variance table based assembly time data

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F	P Value
Fixture(C)	82.80	2	41.40	7.54	0.01
Layout(B)	4.08	1	4.09	0.34	0.58
Operator(Within Layout) B(A)	71.91	6	11.99	5.15	0.01
CA	19.04	2	9.52	1.73	0.22
C× B(A)	65.84	12	5.49	2.36	0.04
Error	56.00	24	2.33		
Total	299.67	47			

Though Table 3 contains balanced data, we shall create room for missing observations by deleting some of the observations and then estimate those missing observations and perform the necessary analysis of variance. Using (14) the estimate of $X_{1111} = 22$ is found to be $M_1 = 24$. Replacing X_{1111} in Table 3 by its estimate and performing the requisite analysis of variance, the results in Table 5 are obtained.

Table 5. Analysis of variance table based on assembly time data with an estimate of one missing value

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F	F_{tab}
Fixture(C)	79.625(77.992)	2	39.813(38.996)	7.503(7.5345)	3.89
Layout(B)	3.000	1	3.000	0.262	5.99
Operator(Within Layout) B(A)	68.583	6	11.431	4.868	2.53
CA	18.375	2	9.188	1.732	3.89
C× B(A)	63.667	12	5.306	2.260	2.20
Error	54.00	23	2.348		
Total	287.250(285.617)	46			

The values inside the brackets are obtained after the adjustment has been made for the bias in the sum of squares due to factor C. The bias B is calculated using Theorem 3. In this regard, $B = 1.633$.

To illustrate the estimation of two missing values using Theorem 2 in a three-factor nested-factorial design, we assume that the values X_{2111} and X_{2121} are missing in Table 3. Their least squares estimates are 24 and 28 respectively. If we ignore data classification according to factor C, we have a two-stage nested-factorial design and the estimates of X_{2111} and X_{2121} can be easily found to be 25.4 and 27.4 respectively. In line with (21), the corrected sum of squares due to factor C is calculated to be Corrected $SS_C = 164.233$.

Table 6: Analysis of variance table based on assembly time data with estimates of two missing values

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F	F_{tab}
Fixture(C)	78.167(164.233)	2	39.083(82.117)	6.90(14.490)	4.75
Layout(B)	4.688	1	4.688	0.380	5.99
Operator(Within Layout) B(A)	74.125	6	12.354	5.126	3.44
CA	19.500	2	9.750	1.720	3.89
C× B(A)	68.00	12	5.667	2.351	2.23
Error	53.000	22	2.410		
Total	297.480(385.546)	45			

It can be deduced from Tables 4, 5 and 6 that the sum of squares due to factor C, error sum of squares and total sum of squares all vary depending on whether the analysis is based on balanced data without estimates of missing values or data with one or more estimates of missing values. Interestingly, the analysis of variance results in the three tables lead to the same conclusion for each of the four sources of variations fixture (C), layout (B), operator (C), B(A), CA and C× B(A), indicating the appropriateness of the missing value estimation technique discussed in this paper.

Conclusion

This study primarily deals with the non-iterative least squares estimation of missing values in a three-factor nested-factorial design. The theoretical results obtained in this paper are predicated on several cases of missing values. In particular, we have paid attention to the cases of one missing value and many missing values in the same cell or different cells. In the case of one missing value, we have shown that the estimate of the missing value is equal to the arithmetic mean of the remaining values in the cell containing the missing value. Similar results are also obtained in the case of many missing values.

In the three-factor nested-factorial design, the factor C is crossed with the other factors. The bias in the sum of squares due to factor C is derived when a missing observation is estimated using the proposed estimator. The bias is shown to be a positive quantity. On the basis of many missing values, an expression for the corrected sum of squares due to factor C is given.

In order to show the application and suitability of the theoretical results, a numerical example based on the data from [2] is considered. Analysis of variance tables are obtained based on the original data, data with the estimate of one missing value and the data with the estimate of two missing values. Correction is also made for the upward bias in the sum of squares due to factor C. Interestingly, the analysis of variance results obtained in these cases lead to the same conclusion for each requisite source of variation.

References

[1] John, P. W. (1998). Statistical design and analysis of experiments. Society for Industrial and Applied Mathematics, Philadelphia.

[2] Montgomery, D. C.(2013). Design and analysis of experiments, eight edition. John wiley ans Sons, USA.

[3] Allan, F. E and Wishart, J. (1930). A method of estimating yield of missing plot in field experimental work. Journal of Agricultural Science, vol. 20, pp. 399-406.

[4] Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. Empire Journal of Experimental Agriculture, 1(2), pp. 129-142.

[5] Bhatra, C. N. Ch and Dharamyadav, T. (2013). Estimation of missing observation in random- ized block design, international Journal of Technology and Engineering Science, vol. 1, no. 6, pp. 618-621.

[6] Shyr, Y and Kshirsagar, A. M. (1997). A formula for a missing plot in a general incomplete block design, when recovery of interblock information is used. Communications in Statistics- Theory and Methods, vol. 26, no. 12, pp. 2855-2860.

[7] Kramer, C. Y and Glass, S. (1960). Analysis of variance of a latin square with missing obser- vations. Journal of the Royal Statistical Society, Series C (Applied Statistics), vol. 9, no. 1, pp. 43-50.

[8] Subramani, J. (1991b). Non-iterative least squares estimation of missing values in Graeco-latin square designs, Biometrical Journal, vol. 33, issue 6, pp. 763-769.

[9] Subramani, J. (1991a). Non-iterative least squares estimation of missing values in replicated latin square designs, Biometrical Journal, vol. 33, issue 8, pp. 999-1011.

[10] Subramani, J and Aggarwal, M. L.(1993). Estimation of several missing values in F-square, Biometrical Journal, vol. 35, issue 4, pp. 455-463.

[11] Subramani, J. (1994). Non-iterative least squares estimation of missing values in cross-over designs without residual effect, Biometrical Journal, vol. 36, issue 3, pp. 285-292.

[12] Ahmed, L. A. (2016). Missing values estimation comparisons in split-plot design. International Journal of Computer and Information Technology, vol. 5, issue 3, pp. 337-344.

[13] Schielzeth, H and Nakagawa, S. (2013). Nested by design: model fitting and interpretation in mixed model era. Methods in Ecology and Evolution, vol. 4, pp. 14-24.

[14] Sahai, H and Ageel, M. I. (2000). The analysis of variance: fixed, random and mixed models. Springer Science +Business Media, New York.

[15] Rangaswamy, R. (2010). A textbook of agricultural statistics, second edition. New Age Inter- national Publishers, New Delhi, India.

[16] Das, M. N and Giri, N. C. (1986). Design and analysis of experiments, Second Edition. New Age International (P) Limited, Publishers, New Delhi, India.