

PAPER • OPEN ACCESS

## Analysis of variance for strip plot design with missing values: bias correction of the mean squares

To cite this article: K Nisa *et al* 2020 *J. Phys.: Conf. Ser.* **1524** 012049

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

# Analysis of variance for strip plot design with missing values: bias correction of the mean squares

K Nisa<sup>1</sup>, N Hamsyiah<sup>1</sup>, M Usman<sup>1</sup>, Warsono<sup>1</sup>

<sup>1</sup>Department of Mathematics, University of Lampung  
Jl. Soemantri Brojonegoro No.1, Bandar Lampung, Indonesia

Corresponding author: [khoirin.nisa@fmipa.unila.ac.id](mailto:khoirin.nisa@fmipa.unila.ac.id)

**Abstract.** Missing values in strip plot design cause an unbalanced design. As a result, several problems occur in analyzing the data. This paper presents the estimation of two and three missing values from strip-plot design with the fixed model using the Yates approach. However, the estimation toward the missing values in strip-plot design causes bias on the treatment mean squares which is shown by its expected value. To omit the bias caused by the estimated values, a procedure of Satterthwaite-Cochran approximation is applied. The procedure corrects the bias and gives an adjusted analysis of variance for the hypothesis test.

## 1. Introduction

In an experimental study, an experiment must be designed following the objective of the research. Experimental design refers to an exact implementation of program planning in the form of an experiment chart and is used to collect data in the study. Although the experimental design has been set up well, practically missing data in experimental results are often found and it is likely to occur in strip-plot design. The missing data in a strip plot design cause several problems in analyzing the data. The problems are raised by the treatment which is not orthogonal toward replication causing some treatments that are not applied to all replication. Therefore, the estimation of missing data in strip-plot design is necessary to be done.

The study of missing data analysis in experimental designs with various approaches has been an interesting topic in statistical research, one can see e.g. [1-6]. A method developed by Yates [6] can be used to estimate those missing data. The Yates' method is simple, i.e. the missing value is estimated based on the available observation; thus, mean square error turns unbiased and remains the same as it is in the gained observations. However, this approach results in the bias of other mean squares parameters. As a consequence, this will lead to biased hypothesis testing. Thus, adjusted testing is urgently required.

In this paper, we present the estimation for three missing data cases in strip plot design with the fixed model using the Yates approach and show the biased treatment mean square resulted from the Yates approach through its expected value. We use the Satterthwaite-Cochran approximation to correct the bias.

Strip plot design refers to experimental design involving two structure treatments, factors A and B, in which the accuracy of their interaction effect is more prominent than the effect of each factor. Factor A is with a level, factor B is with b level, and the experiment unit is classified into r block. The linear model of strip plot design with the basic design of random block design is as follows:



$$Y_{ijk} = \mu + K_k + \tau_i + \vartheta_{ik} + \delta_j + \varphi_{jk} + \gamma_{ij} + \varepsilon_{ijk} ,$$

$$\text{with } i = 1, 2, \dots, a; j = 1, 2, \dots, b; k = 1, 2, \dots, r \quad (1)$$

in which  $Y_{ijk}$  denotes the observation value on  $i$ -th level factor A,  $j$ -th level factor B, and  $k$ -th block;  $\mu$  is the general mean,  $K$  is  $k$ -th block effect,  $\tau_i$  is the effect of the  $i$ -th level of factor A,  $\delta_j$  is the effect of the  $j$ -th level of factor B,  $\gamma_{ij}$  is the interaction effect between the  $i$ -th level of A and the  $j$ -th level of B,  $\vartheta_{ik}$  is the error effect on the  $i$ -th level of A and  $k$ -th block,  $\varphi_{jk}$  is the error effect on  $j$ -th level of B and  $k$ -th block,  $\varepsilon_{ijk}$  is the error effect on  $i$ -th level of A,  $j$ -th level of B, and  $k$ -th block. Assumptions for the linear model (1) are as follow:

1.  $\sum_{i=1}^a \tau_i = \sum_{j=1}^b \delta_j = \sum_{k=1}^r K_k = \sum_{i=1}^a \gamma_{ij} = \sum_{j=1}^b \gamma_{ij} = \sum_{i=1}^a \sum_{j=1}^b \gamma_{ij} = 0$ .
2.  $\vartheta_{ik} \sim N_{iid}(0, \sigma_{\vartheta}^2)$ ,  $\varphi_{jk} \sim N_{iid}(0, \sigma_{\varphi}^2)$ , and  $\varepsilon_{ijk} \sim N_{iid}(0, \sigma_{\varepsilon}^2)$ .

The sum square of each component in the model is obtained using the following formula:

Sum square total (SST):

$$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r Y_{ijk}^2 - \frac{Y_{...}^2}{abr} \quad (2)$$

Sum square of the block (SSK):

$$SSK = \frac{\sum_{k=1}^r Y_{..k}^2}{ab} - \frac{Y_{...}^2}{abr} \quad (3)$$

Sum square of factor A (SSA):

$$SSA = \frac{\sum_{i=1}^a Y_{i..}^2}{br} - \frac{Y_{...}^2}{abr} \quad (4)$$

Sum square error of factor A (SSEa):

$$SSEa = \frac{\sum_{i=1}^a \sum_{k=1}^r Y_{i.k}^2}{b} - \frac{\sum_{i=1}^a Y_{i..}^2}{br} - \frac{\sum_{k=1}^r Y_{..k}^2}{ab} + \frac{Y_{...}^2}{abr} \quad (5)$$

Sum square of factor B (SSB):

$$SSB = \frac{\sum_{j=1}^b Y_{.j.}^2}{ar} - \frac{Y_{...}^2}{abr} \quad (6)$$

Sum square error of factor B (SSEb):

$$SSEb = \frac{\sum_{j=1}^b \sum_{k=1}^r Y_{.jk}^2}{a} - \frac{\sum_{j=1}^b Y_{.j.}^2}{ar} - \frac{\sum_{k=1}^r Y_{..k}^2}{ab} + \frac{Y_{...}^2}{abr} \quad (7)$$

Sum square of interaction between factor A and factor B (SSAB):

$$SSAB = \frac{\sum_{i=1}^a \sum_{j=1}^b Y_{ij.}^2}{r} - \frac{\sum_{i=1}^a Y_{i..}^2}{br} - \frac{\sum_{j=1}^b Y_{.j.}^2}{ar} + \frac{Y_{...}^2}{abr} \quad (8)$$

Sum square error:

$$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r Y_{ijk}^2 - \frac{\sum_{i=1}^a \sum_{j=1}^b Y_{ij.}^2}{r} - \frac{\sum_{i=1}^a \sum_{k=1}^r Y_{i.k}^2}{b} - \frac{\sum_{j=1}^b \sum_{k=1}^r Y_{.jk}^2}{a} + \frac{\sum_{i=1}^a Y_{i..}^2}{br} + \frac{\sum_{j=1}^b Y_{.j.}^2}{ar} + \frac{\sum_{k=1}^r Y_{..k}^2}{ab} - \frac{Y_{...}^2}{abr} \quad (9)$$

The analysis of variance and its expected mean squares for complete data with the fixed model are presented in Table 1.

**Table 1.** Analysis of Variance (ANOVA) for Strip Plot Design

Source of variance	Degree of freedom (Df)	Sum of Squares (SS)	Mean squares (MS) =SS/Df	Expected Mean squares (E(MS) )	F-value
Block	$r - 1$	SSK	MSK	-	-
A	$a - 1$	SSA	MSA	$\sigma_{\varepsilon}^2 + b\sigma_{\theta}^2 + br \frac{\sum_{i=1}^a \tau_i^2}{(a-1)}$	$\frac{MSA}{MSEa}$
Error A	$(a-1)(r-1)$	SSEa	MSEa	$\sigma_{\varepsilon}^2 + b\sigma_{\theta}^2$	-
B	$b - 1$	SSB	MSB	$\sigma_{\varepsilon}^2 + a\sigma_{\phi}^2 + ar \frac{\sum_{j=1}^b \delta_j^2}{(b-1)}$	$\frac{MSB}{MSEb}$
Error B	$(b-1)(r-1)$	SSEb	MSEb	$\sigma_{\varepsilon}^2 + a\sigma_{\phi}^2$	-
AB	$(a-1)(b-1)$	SSAB	MSAB	$\sigma_{\varepsilon}^2 + r \frac{\sum_{i=1}^a \sum_{j=1}^b \gamma_{ij}}{(a-1)(b-1)}$	$\frac{MSAB}{MSE}$
Error	$(a-1)(b-1)(r-1)$	SSE	MSE	$\sigma_{\varepsilon}^2$	-
Total	$abr - 1$	SST			

## 2. Methods

In this paper Yates approach [7] is used to estimate two and three missing data cases in strip plot design. The Yates approach estimator is obtained through mathematical calculation by differentiating partially SSE toward the missing data and equating it by zero, i.e.

$$\frac{\partial SSE}{\partial Y_{ijk}} = 0$$

to obtain  $\hat{Y}_{ijk}$  that minimizing SSE. Without loss of generality, here we set  $a = 3$ ,  $b = 4$  and  $r = 3$  and the two and three missing values are  $\{Y_{111}, Y_{222}\}$  and  $\{Y_{111}, Y_{222}, Y_{333}\}$  respectively. Let  $Y_{ijk}^*$  denotes the values available in the data.

Furthermore, the bias of treatment mean squares caused by the estimation of missing values is obtained by comparing expected mean squares in strip plot model comprising two and three missing data with expected mean squares in complete data. Then, to omit the bias, adjusted variance analysis is done by using Satterthwaite-Cochran approximation [8].

The Satterthwaite-Cochran approximation for linear combinations of means squares can be described as follow. Suppose  $M_1, M_2, \dots, M_k$  are independent mean squares and that

$$\frac{d_i M_i}{E(M_i)} \sim \chi_{d_i}^2, \quad \forall i = 1, 2, \dots, k.$$

It follows that

$$E\left(\frac{d_i M_i}{E(M_i)}\right) = d_i, \quad \text{var}\left(\frac{d_i M_i}{E(M_i)}\right) = 2d_i, \quad \text{and} \quad M_i \sim \frac{E(M_i)}{d_i} \chi_{d_i}^2$$

for all  $i = 1, 2, \dots, k$ . The Satterthwaite-Cochran approximation works by assuming that  $M$  is approximately distributed as a scaled  $\chi^2$  like each of the variables in the linear combination above. One can see e.g. [9]-[10] for more details on the approximation.

### 3. Result and discussion

#### 3.1. Estimation result

For two missing values  $Y_{111}$  and  $Y_{222}$  the SSE in (9) can be rewritten as follow

$$\begin{aligned}
 SSE = & Y_{111}^2 + Y_{222}^2 + \sum_{i=3}^a \sum_{j=3}^b \sum_{k=1}^r Y_{ijk}^2 - \frac{1}{r} \left[ (Y_{111} + Y_{11..}^*)^2 + (Y_{222} + Y_{22..}^*)^2 + \sum_{i=3}^a \sum_{j=3}^b Y_{ij.}^2 \right] \\
 & - \frac{1}{b} \left[ (Y_{111} + Y_{1.1}^*)^2 + (Y_{222} + Y_{2.2}^*)^2 + \sum_{i=3}^a \sum_{k=3}^r Y_{i.k}^2 \right] \\
 & - \frac{1}{a} \left[ (Y_{111} + Y_{.11}^*)^2 + (Y_{222} + Y_{.22}^*)^2 + \sum_{j=3}^b \sum_{k=3}^r Y_{.jk}^2 \right] \\
 & + \frac{1}{br} \left[ (Y_{111} + Y_{1..}^*)^2 + (Y_{222} + Y_{2..}^*)^2 + \sum_{i=3}^a Y_{i..}^2 \right] + \frac{1}{ar} \left[ (Y_{111} + Y_{.1.}^*)^2 + (Y_{222} + Y_{.2.}^*)^2 + \sum_{j=3}^b Y_{.j.}^2 \right] \\
 & + \frac{1}{ab} \left[ (Y_{111} + Y_{.1.}^*)^2 + (Y_{222} + Y_{.2.}^*)^2 + \sum_{k=3}^r Y_{..k}^2 \right] \\
 & - \frac{1}{abr} (Y_{111} + Y_{222} + Y_{...}^*)^2 \quad (10)
 \end{aligned}$$

The first derivatives of (10) with respect to  $Y_{111}$  and  $Y_{222}$  give respectively;

$$\begin{aligned}
 \frac{\partial SSE}{\partial Y_{111}} = & 2Y_{111} - \frac{2}{r} (Y_{111} + Y_{11..}^*) - \frac{2}{b} (Y_{111} + Y_{1.1}^*) - \frac{2}{a} (Y_{111} + Y_{.11}^*) + \frac{2}{br} (Y_{111} + Y_{1..}^*) + \\
 & \frac{2}{ar} (Y_{111} + Y_{.1.}^*) + \frac{2}{ab} (Y_{111} + Y_{.1.}^*) - \frac{2}{abr} (Y_{111} + Y_{222} + Y_{...}^*) \quad (11)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial SSE}{\partial Y_{222}} = & 2Y_{222} - \frac{2}{r} (Y_{222} + Y_{22..}^*) - \frac{2}{b} (Y_{222} + Y_{2.2}^*) - \frac{2}{a} (Y_{222} + Y_{.22}^*) + \frac{2}{br} (Y_{222} + Y_{2..}^*) + \\
 & \frac{2}{ar} (Y_{222} + Y_{.2.}^*) + \frac{2}{ab} (Y_{222} + Y_{.2.}^*) - \frac{2}{abr} (Y_{111} + Y_{222} + Y_{...}^*) \quad (12)
 \end{aligned}$$

Equalizing (11) and (12) with zero for minimizing SSE we obtain:

$$\begin{aligned}
 \hat{Y}_{111} = & [(a-1)(b-1)(r-1)(abY_{11..}^* + arY_{1.1}^* + brY_{.11}^* - aY_{1..}^* - bY_{.1.}^* - rY_{.1.}^*) + abY_{22..}^* \\
 & + arY_{2.2}^* + brY_{.22}^* - aY_{2..}^* - bY_{.2.}^* - rY_{.2.}^* + \{(a-1)(b-1)(r-1) + 1\}Y_{...}^*] \\
 & / [(a-1)(b-1)(r-1) - 1][(a-1)(b-1)(r-1) + 1] \quad (13)
 \end{aligned}$$

and

$$\begin{aligned}
 \hat{Y}_{222} = & [(a-1)(b-1)(r-1)(abY_{22..}^* + arY_{2.2}^* + brY_{.22}^* - aY_{2..}^* - bY_{.2.}^* - rY_{.2.}^*) + abY_{11..}^* \\
 & + arY_{1.1}^* + brY_{.11}^* - aY_{1..}^* - bY_{.1.}^* - rY_{.1.}^* + \{(a-1)(b-1)(r-1) + 1\}Y_{...}^*] \\
 & / [(a-1)(b-1)(r-1) - 1][(a-1)(b-1)(r-1) + 1] \quad (14)
 \end{aligned}$$

For three missing values  $Y_{111}$ ,  $Y_{222}$  and  $Y_{333}$ , we rewrite the SSE as follows

$$\begin{aligned}
 SSE = & Y_{111}^2 + Y_{222}^2 + Y_{333}^2 + \sum \sum \sum Y_{ijk}^2 \\
 & - \frac{1}{r} \left[ (Y_{111} + Y_{11..})^2 + (Y_{222} + Y_{22..})^2 + (Y_{333} + Y_{33..})^2 + \sum_{i=4}^a \sum_{j=4}^b Y_{ij.}^2 \right] \\
 & - \frac{1}{b} \left[ (Y_{111} + Y_{1.1})^2 + (Y_{222} + Y_{2.2})^2 + (Y_{333} + Y_{3.3})^2 + \sum_{i=4}^a \sum_{k=4}^r Y_{i.k}^2 \right] \\
 & - \frac{1}{a} \left[ (Y_{111} + Y_{11.})^2 + (Y_{222} + Y_{22.})^2 + (Y_{333} + Y_{33.})^2 + \sum_{j=4}^b \sum_{k=4}^r Y_{.jk}^2 \right] \\
 & + \frac{1}{br} \left[ (Y_{111} + Y_{1..})^2 + (Y_{222} + Y_{2..})^2 + (Y_{333} + Y_{3..})^2 + \sum_{i=4}^a Y_{i..}^2 \right] \\
 & + \frac{1}{ar} \left[ (Y_{111} + Y_{1.})^2 + (Y_{222} + Y_{2.})^2 + (Y_{333} + Y_{3.})^2 + \sum_{j=4}^b Y_{.j.}^2 \right] \\
 & + \frac{1}{ab} \left[ (Y_{111} + Y_{1.1})^2 + (Y_{222} + Y_{2.2})^2 + (Y_{333} + Y_{3.3})^2 + \sum_{k=4}^r Y_{..k}^2 \right] \\
 & - \frac{1}{abr} (Y_{111} + Y_{222} + Y_{333} + Y_{...})^2 \tag{15}
 \end{aligned}$$

The first derivatives of SSE in (15) with respect to  $Y_{111}$ ,  $Y_{222}$  and  $Y_{333}$  give respectively;

$$\begin{aligned}
 \frac{\partial JKG}{\partial Y_{111}} = & 2Y_{111} - \frac{2}{r}(Y_{111} + Y_{11..}) - \frac{2}{b}(Y_{111} + Y_{1.1}) - \frac{2}{a}(Y_{111} + Y_{11.}) + \frac{2}{br}(Y_{111} + Y_{1..}) \\
 & + \frac{2}{ar}(Y_{111} + Y_{1.}) + \frac{2}{ab}(Y_{111} + Y_{1.1}) - \frac{2}{abr}(Y_{111} + Y_{222} + Y_{333} + Y_{...}) \tag{16}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial JKG}{\partial Y_{222}} = & 2Y_{222} - \frac{2}{r}(Y_{222} + Y_{22..}) - \frac{2}{b}(Y_{222} + Y_{2.2}) - \frac{2}{a}(Y_{222} + Y_{22.}) + \frac{2}{br}(Y_{222} + Y_{2..}) \\
 & + \frac{2}{ar}(Y_{222} + Y_{2.}) + \frac{2}{ab}(Y_{222} + Y_{2.2}) - \frac{2}{abr}(Y_{111} + Y_{222} + Y_{333} + Y_{...}) \tag{17}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial JKG}{\partial Y_{333}} = & 2Y_{333} - \frac{2}{r}(Y_{333} + Y_{33..}) - \frac{2}{b}(Y_{333} + Y_{3.3}) - \frac{2}{a}(Y_{333} + Y_{33.}) + \frac{2}{br}(Y_{333} + Y_{3..}) \\
 & + \frac{2}{ar}(Y_{333} + Y_{3.}) + \frac{2}{ab}(Y_{333} + Y_{3.3}) - \frac{2}{abr}(Y_{111} + Y_{222} + Y_{333} + Y_{...}) \tag{18}
 \end{aligned}$$

Applying  $\frac{\partial SSE}{\partial Y_{ijk}} = 0$  to (16), (17) and (18) gives the three estimates of the missing values as follow

$$\begin{aligned}
SSE = & Y_{111}^2 + Y_{222}^2 + Y_{333}^2 + \sum \sum \sum Y_{ijk}^2 \\
& - \frac{1}{r} \left[ (Y_{111} + Y_{11.})^2 + (Y_{222} + Y_{22.})^2 + (Y_{333} + Y_{33.})^2 + \sum_{i=4}^a \sum_{j=4}^b Y_{ij.}^2 \right] \\
& - \frac{1}{b} \left[ (Y_{111} + Y_{1.1})^2 + (Y_{222} + Y_{2.2})^2 + (Y_{333} + Y_{3.3})^2 + \sum_{i=4}^a \sum_{k=4}^r Y_{i.k}^2 \right] \\
& - \frac{1}{a} \left[ (Y_{111} + Y_{.11})^2 + (Y_{222} + Y_{.22})^2 + (Y_{333} + Y_{.33})^2 + \sum_{j=4}^b \sum_{k=4}^r Y_{.jk}^2 \right] \\
& + \frac{1}{br} \left[ (Y_{111} + Y_{1..})^2 + (Y_{222} + Y_{2..})^2 + (Y_{333} + Y_{3..})^2 + \sum_{i=4}^a Y_{i..}^2 \right] \\
& + \frac{1}{ar} \left[ (Y_{111} + Y_{.1.})^2 + (Y_{222} + Y_{.2.})^2 + (Y_{333} + Y_{.3.})^2 + \sum_{j=4}^b Y_{.j.}^2 \right] \\
& + \frac{1}{ab} \left[ (Y_{111} + Y_{.1.})^2 + (Y_{222} + Y_{.2.})^2 + (Y_{333} + Y_{.3.})^2 + \sum_{k=4}^r Y_{..k}^2 \right] \\
& - \frac{1}{abr} (Y_{111} + Y_{222} + Y_{333} + Y_{...})^2 \tag{19}
\end{aligned}$$

The first derivatives of SSE in (15) with respect to  $Y_{111}$ ,  $Y_{222}$  and  $Y_{333}$  give respectively;

$$\begin{aligned}
\frac{\partial JKG}{\partial Y_{111}} = & 2Y_{111} - \frac{2}{r}(Y_{111} + Y_{11.}) - \frac{2}{b}(Y_{111} + Y_{1.1}) - \frac{2}{a}(Y_{111} + Y_{.11}) + \frac{2}{br}(Y_{111} + Y_{1..}) \\
& + \frac{2}{ar}(Y_{111} + Y_{.1.}) + \frac{2}{ab}(Y_{111} + Y_{.1.}) - \frac{2}{abr}(Y_{111} + Y_{222} + Y_{333} + Y_{...}) \tag{20}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial JKG}{\partial Y_{222}} = & 2Y_{222} - \frac{2}{r}(Y_{222} + Y_{22.}) - \frac{2}{b}(Y_{222} + Y_{2.2}) - \frac{2}{a}(Y_{222} + Y_{.22}) + \frac{2}{br}(Y_{222} + Y_{2..}) \\
& + \frac{2}{ar}(Y_{222} + Y_{.2.}) + \frac{2}{ab}(Y_{222} + Y_{.2.}) - \frac{2}{abr}(Y_{111} + Y_{222} + Y_{333} + Y_{...}) \tag{21}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial JKG}{\partial Y_{333}} = & 2Y_{333} - \frac{2}{r}(Y_{333} + Y_{33.}) - \frac{2}{b}(Y_{333} + Y_{3.3}) - \frac{2}{a}(Y_{333} + Y_{.33}) + \frac{2}{br}(Y_{333} + Y_{3..}) \\
& + \frac{2}{ar}(Y_{333} + Y_{.3.}) + \frac{2}{ab}(Y_{333} + Y_{.3.}) - \frac{2}{abr}(Y_{111} + Y_{222} + Y_{333} + Y_{...}) \tag{22}
\end{aligned}$$

Applying  $\frac{\partial SSE}{\partial Y_{ijk}} = 0$  to (16), (17) and (18) gives the three estimates of the missing values as follow

$$\begin{aligned}
\hat{Y}_{111} = & [\{(a-1)(b-1)(r-1) - 1\}(abY_{11.} + arY_{1.1} + brY_{.11} - aY_{1..} - bY_{.1.} - rY_{.1.}) \\
& + ab(Y_{22.} + Y_{33.}) + ar(Y_{2.2} + Y_{3.3}) + br(Y_{.22} + Y_{.33}) - a(Y_{2..} + Y_{3..}) \\
& - b(Y_{.2.} + Y_{.3.}) - r(Y_{.2.} + Y_{.3.}) + \{(a-1)(b-1)(r-1) + 1\}Y_{...}] \\
& / [\{(a-1)(b-1)(r-1) + 1\}\{(a-1)(b-1)(r-1) - 2\}] \tag{23}
\end{aligned}$$

$$\hat{Y}_{222} = [\{(a-1)(b-1)(r-1) - 1\} (abY_{22..}^* + arY_{2.2.}^* + brY_{.22.}^* - aY_{2..}^* - bY_{.2.}^* - rY_{..2}^*) + ab(Y_{11..}^* + Y_{33..}^*) + ar(Y_{1.1.}^* + Y_{3.3.}^*) + br(Y_{.11.}^* + Y_{.33.}^*) - a(Y_{1..}^* + Y_{3..}^*) - b(Y_{.1.}^* + Y_{.3.}^*) - r(Y_{..1}^* + Y_{..3}^*) + \{(a-1)(b-1)(r-1) + 1\}Y_{...}^*] / [\{(a-1)(b-1)(r-1) + 1\}\{(a-1)(b-1)(r-1) - 2\}] \quad (24)$$

$$\hat{Y}_{333} = [\{(a-1)(b-1)(r-1) - 1\} (abY_{33..}^* + arY_{3.3.}^* + brY_{.33.}^* - aY_{3..}^* - bY_{.3.}^* - rY_{..3}^*) + ab(Y_{11..}^* + Y_{22..}^*) + ar(Y_{1.1.}^* + Y_{2.2.}^*) + br(Y_{.11.}^* + Y_{.22.}^*) - a(Y_{1..}^* + Y_{2..}^*) - b(Y_{.1.}^* + Y_{.2.}^*) - r(Y_{..1}^* + Y_{..2}^*) + \{(a-1)(b-1)(r-1) + 1\}Y_{...}^*] / [\{(a-1)(b-1)(r-1) + 1\}\{(a-1)(b-1)(r-1) - 2\}] \quad (21)$$

### 3.2. Expected Mean Squares

Below we provide the expected mean squares for complete data with fixed model (1) for the case of  $a = 3$ ,  $b = 4$  and  $r = 3$ .

**Table 2.** The expected mean squares for complete data with  $a = 3$ ,  $b = 4$  and  $r = 3$ .

Source of Variance	df	MS	E(MS)
Block	2	MSK	-
A	2	MSA	$\sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2 + 6 \sum_{i=1}^3 \tau_i^2$
Error A	4	MSEa	$\sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2$
B	3	MSB	$\sigma_{\varepsilon}^2 + 3\sigma_{\varphi}^2 + 3 \sum_{j=1}^4 \delta_j^2$
Error B	6	MSEb	$\sigma_{\varepsilon}^2 + 3\sigma_{\varphi}^2$
AB	6	MSAB	$\sigma_{\varepsilon}^2 + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^4 \gamma_{ij}^2$
Error	12	MSE	$\sigma_{\varepsilon}^2$

When there are missing values in strip plot design, the expected mean squares is obtained by substituting the linear model for the estimated values and the linear model for the sum of the available observations into Equation (2) to (9), then divided by its degree of freedom, i.e. the degree of freedom for total error subtracted by the number of missing values. Then, the expected mean squares are found. We successfully obtained the expected mean squares for two and three missing values cases and we present the summary in Table 3 and Table 4.

**Table 3.** The expected mean squares for data with two missing values

Source of Variance	df	MS	E(MS)
Block	2	MSK	-
A	2	MSA	$\frac{166}{143} \sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2 + 6 \sum_{i=1}^3 \tau_i^2$
Error (a)	4	MSEa	$\frac{335}{286} \sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2$



B	3	MSB	$\frac{499}{429}\sigma_{\varepsilon}^2 + 3\sigma_{\varphi}^2 + 3\sum_{j=1}^4\delta_j^2$
Error (b)	6	MSEb	$\frac{502}{429}\sigma_{\varepsilon}^2 + 3\sigma_{\varphi}^2$
AB	6	MSAB	$\frac{502}{429}\sigma_{\varepsilon}^2 + \frac{1}{2}\sum_{i=1}^3\sum_{j=1}^4\gamma_{ij}^2$
Error	10	MSE	$\sigma_{\varepsilon}^2$

**Table 4.** The expected mean squares for data with three missing values

Source of Variance	Df	MS	E(MS)
Block	2	MSK	-
A	2	MSA	$\frac{16}{13}\sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2 + 6\sum_{i=1}^3\tau_i^2$
Error A	4	MSEa	$\frac{329}{260}\sigma_{\theta}^2 + 4\sigma_{\theta}^2$
B	3	MSB	$\frac{161}{130}\sigma_{\varepsilon}^2 + 3\sigma_{\varphi}^2 + 3\sum_{j=1}^4\delta_j^2$
Error B	6	MSEb	$\frac{82}{65}\sigma_{\varepsilon}^2 + 3\sigma_{\varphi}^2$
AB	6	MSA B	$\frac{82}{65}\sigma_{\varepsilon}^2 + \frac{1}{2}\sum_{i=1}^3\sum_{j=1}^4\gamma_{ij}^2$
Error	9	MSE	$\sigma_{\varepsilon}^2$

In Table 3 and Table 4 the expected mean square error of data with missing values gives the same result as the expected mean square error of complete data in Table 2, namely  $\sigma_{\varepsilon}^2$ . Therefore, it is revealed that the estimation of missing values using Yates approach gives an unbiased mean square error. However, there are biases in treatment mean squares. In our cases here, the bias occurs in MSA, MSB and MSAB. The biases for two and three missing values data are summarized in Table 5.

**Table 5.** The bias of the mean squares

MS	Two missing values	Three missing values
MSA	$23/143 \approx 0,161$	$3/13 \approx 0,231$
MSB	$70/429 \approx 0,163$	$31/130 \approx 0,238$
MSAB	$71/429 \approx 0,165$	$17/65 \approx 0,262$

In order to omit the bias, the Satterthwaite-Cochran approximation is applied and described in the following section.

### 3.3. Bias correction of mean squares

Following Bancroft [11], the biases of the treatment mean squares are corrected by using Satterthwaite-Cochran approximation. The procedures for the two missing values case are as follows;

i. For factor A,

$$MSA_{adj} = MSA - \frac{23}{143}MSE,$$

and

$$MSEa_{adj} = MSEa - \frac{49}{286}MSE.$$

Thus, the expected values of adjusted mean squares factor A and adjusted mean square error factor A are obtained as follows

$$E(MSA_{adj}) = \sigma_{\epsilon}^2 + 4\sigma_{\theta}^2 + 6 \sum_{i=1}^3 \tau_i^2$$

and

$$E(MSEa_{adj}) = \sigma_{\epsilon}^2 + 4\sigma_{\theta}^2.$$

Then the F-test statistics for factor is:

$$F' = \frac{MSA_{adj}}{MSEa_{adj}},$$

with degree of freedom for the numerator and denominator are 2 and 4 respectively.

ii. For factor B,

$$MSB_{adj} = MSB - \frac{70}{429}MSE,$$

and

$$MSEb_{adj} = MSEb - \frac{73}{429}MSE.$$

Giving expected values:

$$E(MSEa_{adj}) = \sigma_{\epsilon}^2 + 3\sigma_{\phi}^2 + 3 \sum_{j=1}^4 \delta_j^2$$

and

$$E(MSEb_{adj}) = \sigma_{\epsilon}^2 + 3\sigma_{\phi}^2.$$

Then the F-test statistics for factor A is:

$$F' = \frac{MSB_{adj}}{MSEb_{adj}},$$

with degree of freedom for the numerator and denominator are 3 and 6 respectively.

iii. For interaction between factor A and factor B,

$$MSAB_{adj} = cMSAB$$

with  $c = \frac{429}{502}$ , then the expected values of MSAB is:

$$E(MSAB_{adj}) = \sigma_{\epsilon}^2 + \psi_2$$

where  $\psi_2 = \frac{429}{1004} \sum_{i=1}^3 \sum_{j=1}^4 \gamma_{ij}^2$ . Thus the F-test statistics for interaction between factor A and B is:

$$F' = \frac{MSEAB_{adj}}{MSE},$$

with degree of freedom for the numerator and denominator are 6 and 9 respectively.

Now we have the expected values of adjusted mean squares are the same as the expected mean squares for complete data in strip plot design. For the case of three missing values, the adjusted mean squares for each treatment are obtained using the same procedure above, i.e.

i. For factor A,

$$MSA_{adj} = MSA - \frac{3}{13}MSE,$$

and

$$MSEa_{adj} = MSEa - \frac{69}{260}MSE.$$

ii. For factor B,

$$MSB_{adj} = MSB - \frac{31}{130}MSE,$$

and

$$MSEb_{adj} = MSEb - \frac{17}{65}MSE.$$

iii. For interaction between factor A and factor B,

$$MSAB_{adj} = cMSAB, \text{ with } c = \frac{65}{82}.$$

We summarize the adjusted analysis of variance for the cases in Table 6.

**Table 6.** Adjusted Analysis of Variance using Satterthwaite-Cochran Approximation

Source of Variance	df	Mean squares		E(MS)
		Two missing values	Three missing values	
Block	2	MSK	MSK	-
A	2	$MSA - \frac{23}{143}MSE$	$MSA - \frac{3}{13}MSE$	$\sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2 + 6 \sum_{i=1}^3 \tau_i^2$
Error A	4	$MSEa - \frac{49}{286}MSE$	$MSEa - \frac{69}{260}MSE$	$\sigma_{\varepsilon}^2 + 4\sigma_{\theta}^2$
B	3	$MSB - \frac{70}{429}MSE$	$MSB - \frac{70}{429}MSE$	$\sigma_{\varepsilon}^2 + 3\sigma_{\phi}^2 + 3 \sum_{j=1}^4 \delta_j^2$
Error B	6	$MSEb - \frac{73}{429}MSE$	$MSEb - \frac{73}{429}MSE$	$\sigma_{\varepsilon}^2 + 3\sigma_{\phi}^2$
AB	6	$\frac{429}{502}MSAB$	$\frac{429}{502}MSAB$	$\sigma_{\varepsilon}^2 + \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^4 \gamma_{ij}^2$
Error	12	MSE	MSE	$\sigma_{\varepsilon}^2$

#### 4. Conclusion

In this paper we show that the estimation toward several missing values in strip plot design using Yate's approach gives biased treatment mean squares. Then we described the bias correction using Satterthwaite-Cochran approximation and it is shown that the procedure successfully omits the bias.

**References**

- [1] Qumsiyeh M and Kirchner K 2011 *Journal of Statistical Theory and Application* **5** (2) 131-147
- [2] Agbehadji I E, Millham R C, Fong S J and Yang H 2018 *Bioinspired Computational Approach to Missing Value Estimation* Mathematics Problem in Engineering 1-16
- [3] Noble D W A and Nakagawa S 2018 Planned missing data design: stronger inferences, increased research efficiency and improved animal welfare in ecology and evolution. *BioRxiv preprint* [Online]. Available: <https://www.biorxiv.org/content/early/2018/01/11/247064>. [Accessed: Nov. 12, 2018]
- [4] Azadeh A, Saberi Z, Behrouznia H, Radmehr F and Pazhoheshfar P 2012 *Estimating Missing Values in Analysis of Variance Tables by a Flexible Adaptive Artificial Neural Network and Fuzzy Regression Models* Presented in the 6th International Days of Statistics and Economics Prague
- [5] Ahmed L A 2016 *International Journal of Computer and Information Technology* **5** (3) 337-344
- [6] Usman M and Warsono 2003 *Proceedings of Conference of Statistical and Mathematical science of Islamic Society in South East Asia Region* Universitas Islam Bandung
- [7] Yates F 1933 *Empire Journal of Experimental Agriculture* **1** (2) 129-142
- [8] Cochran W G 1977 *Sampling Techniques* Third Edition John Wiley & Sons Inc. New York
- [9] Satterthwaite F E 1964 *An Approximate Distribution of Estimates of Variance Components*. *Biometrics* **2** (6) 110-114
- [10] Davenport J and Webster J T 1972 *Technometrics* **14** (3) 555-569
- [11] Bancroft T A 1968 *Topics in Intermediate Statistical Methods* Vol. 1 Iowa State University Press Ames