RASHI 3 (2) : 50 - 56 (2018) ISSN : 2581 9755

Comperative study of different analysis models for missing observations in multi-factor experiments

Ankita Dutta, Parvez Mallick and Anurup Majumder

Bidhan Chandra Krishi Viswavidyalaya, P.O. Krishi Viswavidyalaya, Mohanpur-741252, Nadia, West Bengal, India.

Received : 06-02-2019 ; Revised : 06-04-2019 ; Accepted : 15-04-2019

ABSTRACT

The study mainly aims at estimation of missing observations and the data analysis in multi-factor experiments through ANCOVA and ANOVA models. The mentioned models are compared on the basis of field experiments on jute in a two factor factorial design in randomized block design lay-out. The study has included the experiments with single missing observation and two missing observations. The field trials were conducted on jute varieties (1st factor) under different fertilizer management schedules (2nd factor) at Mondouri Teaching Farm of Bidhan Chandra Krishi Viswavidyalaya, West Bengal for consecutive two years (2013-14 and 2014-15). Two different models, viz., ANCOVA (Coons, 1957) and ANOVA (Yates, 1933) are compared on the basis of estimated missing values and error mean square values for single missing observation and two missing observations in Factorial RBD set up. It is observed that for a single missing observation and two missing observations in Factorial RBD design, for both of the above models, the estimated values are same. But the error mean squares of ANCOVA model are smaller than those for ANOVA model for experiments of both years.

***Keywords*:** ANCOVA model; ANOVA model; Factorial experiment in RBD set-up.

1. INTRODUCTION

Experimental data with missing observations are very common, specially, in field experiments. The observations may be lost or may be affected by some extraneous causes that cannot be used in analysis procedure. Such data with missing observation(s) are generally analysed through the technique of missing plot. In the present study, the models, *viz*., ANCOVA and ANOVA are used to estimate the single and double missing observations in factorial experiment in RBD set-up. The method of minimising the error sum of squares by using ordinary least square technique (ANOVA) has been used widely in available literature. It has been established (Anderson; 1946) that the sum of squares may be over estimated by the above method. Alternatively, use of the technique of ANCOVA for missing data analysis may provide a better solution. The methodologies of ANCOVA for single missing value or multiple missing values are available in several statistical books and articles. But the application of ANCOVA for missing data analysis in multifactor experiments is seldom used in practice.

Available literature survey reveals that the analysis of missing observation (or observations) has been discussed by several statisticians, since early half of 20th century. Estimation of missing yield was introduced by Allan & Wishart (1930). Yates (1933) estimated the missing observations which minimized the residual sum of square, in addition he also obtained the correct least squares estimates of all estimable parameters. Actually, the method developed by Yates (1933) was an extension of Allan and Wishart (1930) from single missing observation to multiple missing observations in a randomized block design or in a latin square design. Coons (1957) vividly discussed the application of ANCOVA model for estimation of missing observation or observations in multifactor experiments. However, Anderson (1947), Bartlett (1937), *etc*., also worked with ANCOVA model to estimate missing observations earlier to Coons (1957).

The aim of the present study is to compare the methods i.e. Analysis of Covariance (ANCOVA) and method of error minimisation through ordinary least square technique or ANOVA for estimating the single missing observation and double missing observations in factorial experiment under RBD set-up. The methods are compared on the basis of absolute difference (AE) and error mean square (MSE). The bias in treatment sum of square due to missing value estimation is also calculated.

2. MATERIALS AND METHODS

2.1 : Estimation of missing observations in factorial experiment in RBD set-up with ANCOVA (Coons, 1957)

ANCOVA can be used to estimate the missing values in an experiment. The procedure of estimating of missing observations in factorial RBD through ANCOVA is similar to the general ANCOVA analysis except the assigned

E-mail : ankita.manoj1993@gmail.com

values of covariate variables under study. Coons (1957) suggested the procedure to estimate the missing values in factorial experiments through ANCOVA model and the steps are:

For single missing observation :

I. To estimate the single missing observation of response variable, set Y = 0 and assign one covariate.

II. Assign the value of covariate as X = -n (n= total no. of observation) for the experimental unit with the missing observation, and X = 0 otherwise.

III. With the complete set of data for the Y variable and the X variable as assigned above, perform the ANCOVA following the standard procedure.

IV. Estimate the missing observation by using following formula,

$$\hat{y} = -nb_{yx} - nE_{yx} / E_{xx}$$

where, b_{yx} is regression co-efficient of y on x.

For two missing observation :

I. Follow the above mentioned procedure but with two covariates as two observations are missing from response variable.

II. From one response variable and two covariates, two simultaneous equations can be developed, i.e.,

$$E_{X_1X_1}b_1 + E_{X_1X_2}b_2 = E_{X_1}Y$$

$$E_{X_1X_2}b_1 + E_{X_2X_2}b_2 = E_{X_2}Y$$

where, is regression co-efficient of y on x_1 and is regression co-efficient of y on x_2.

By solving these equations we can get b_1 and b_2.

III. To estimate the missing values, the following formula can be used:

$\hat{y}_1 = -nb_1$ and $\hat{y}_2 = -nb_2$

where, $\hat{y}_1$ and $\hat{y}_2$ are the two missing values in response variable.

2.2: Estimation of missing observation through the method of OLS (ANOVA) in a factorial design in RBD set-up (Yates, 1933)

I. For single missing observation:

In case of factorial RBD with **r** replications and two factors having **a** & **b** levels, respectively, if a single observation under study is missing, then the estimated missing value will be,

$$\hat{x} = \frac{rB_{i'} + abT_{j'k'} - G'}{(r-1)(ab-1)} \qquad (1)$$

where,

$\hat{x}$	be the value of the missing observation
r	be the number of replication
ab	be the total number of treatment combinations
$B_{i'}$	be the total of available observations in the i´-th block (with the missing observation) (i= 1,2,..., r)
$T_{j'k'}$	be the total of available observations for the j´k´ -th treatment combination (for which the observation is missing) (j = 1, 2,..., m ; k = 1, 2,..., n)
G'	be the grand total of all the available observations.

Table 2.1: ANOVA for estimating the single missing observation in factorial design in RBD set up:

(S.O.V.)	(d.f.)	(S.S.)	(M.S.)	(F.Cal)
Replication	r-1	RSS	$\frac{RSS}{r-1}$	$F_{cal} = \frac{RMS}{ErMS}$
Treatments:			$\frac{SSA}{a-1}$	$F_{cal} = \frac{TrMS}{ErMS}$
Factor A	a-1	SSA	$\frac{SSB}{b-1}$	
Factor B	b-1	SSB	$\frac{SSAB}{(a-1)(b-1)}$	
Interaction (A×B)	(a-1)(b-1)	SSAB		
Error	(r-1)(ab-1)-1	ErSS	$\frac{ErSS}{(r-1)(ab-1)-1}$	
Treatment (Adjusted)	ab-1	Adj. TrSS	$\frac{Adj.TrSS}{ab-1}$	$F_{cal} = \frac{Adj.TrMS}{Adj.ErMS}$
Error (Adjusted)	(r-1)(ab-1)-1	Adj. ErSS	$\frac{Adj.ErSS}{(r-1)(ab-1)-1}$	

We can get the adjusted treatment sum of square by subtracting the bias from the over-estimated treatment sum of square. Yates (1933) proved that in a complete analysis of variance using the above estimated missing-plot value, the treatment sum of square is over-estimated. According to him, the over estimation may be corrected by subtracting the amount of bias. The amount of bias in a randomized block experiment with one missing unit was given as:

$$\frac{1}{v(v-1)}[B-(v-1)y]^2 \qquad (2)$$

II. For two missing observations :

In case of two missing observations under the above mentioned situation, the same calculation is done in iterative manner to estimate the missing observations. At first, one missing observation is assumed to be the treatment mean value for that missing observation then following the above procedure to estimate the another missing observation. After getting the second estimated value, the first missing observation will be estimated by following the above estimation procedure. The process is then repeated till the same values will come, in each step.

For two missing observations, there is no such formula to estimate the bias in treatment sum of square. So, to estimate the bias, at first the analysis of variance is done with the two estimated missing values having error degrees of freedom less two than the usual analysis procedure. Then the same analysis is done without the estimated missing values but the error sum of square is same from the previous analysis. Thus we can get the treatment sum of square by subtracting the block sum of square and the error sum of square from the total sum of square. Now the difference between the two treatment sum of squares is the amount of bias in treatment sum of square which is estimated due to the iterative method of estimation of missing values.

2.3 : Statistical measures for comparison of the methods (2.1 and 2.2)

After the missing value estimation, using the estimated values the usual computational procedures of the analysis of variance is applied to the augmented data set with some modifications i.e. subtract one from the error degree of freedom for each missing value. Thereafter two models, *i.e.*, ANCOVA and ANOVA are compared on the basis of estimated values and error mean square (MSE).

Mean Square Error is nothing but the ratio of the error sum of square to its degree of freedom.

2.4: Experimental details

The data for factorial design in RBD set-up were collected from the experiment conducted on 'Performance of Jute varieties (*Corchorus capsularis)* under different fertilizer management schedules' by Mrs. Parveen Zaman for her Master's dissertation work under the guidance of Dr. P. Bandopadhyay, Professor, Department of Agronomy, BCKV. The experiment was conducted at Mondouri Teaching Farm of Bidhan Chandra Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal for two consecutive years (2013 and 2014). This farm is located very close to the Tropic of Cancer having approximately 89 ^{0}E longitudes, 23 ^{0}N latitude and about 9.75 in altitude above the sea level. The experimental site was of medium land with irrigation facilities from deep tube well. The soil was neutral and clay loam, with 7.5 pH, 0.56% organic carbon, 220, 57 and 190 kg/ha of available N, P and K respectively. The treatments under study are varieties and nutrient schedules-

Table 2.2: The experimental details for factorial RBD set-up

First factor Treatment: Variety (V)	Second factor Treatment: Nutrient schedule (N)
V1: JRC 321	N1: 60:13:25 Kg ha^{-1}
V2: JRC 698	N2: 80:17.5:33.3 Kg ha^{-1}
V3: JRCJ – 2	N3: 100:21.8:41.7 Kg ha^{-1}

The experiment was conducted in three randomised blocks. The fibre yield of jute crop in q ha^{-1} was taken from the above said experiment for this study.

3. RESULTS AND DISCUSSION

We studied two different situations for estimation of missing observations in case of factorial RBD experiment, *i.e.* a) single missing observation, and b) two missing observations.

Here two different methods are applied to estimate the values of missing observations for each situation. The results from the experiments described in section 2.4 are presented for two consecutive years (2013-14 and 2014-15).

3.1: Estimation of single missing observation

The results of the experiment described in section 2.4 were recorded for two consecutive years (2013-14 and 2014-15). The study analysed estimation of missing value by ANCOVA model and the ANOVA model, when single observation is missing in factorial design in RBD set-up. Table 3.1.1 represents the estimated missing values of different positions of the experiment for the year 2013-14. The table also presents the bias in sum of squares which were over estimated as mentioned in section 2.2.

Table 3.1.1: Estimated value and bias in SS of the treatment for 2013-14

Position	Estimated value by ANCOVA	Estimated value by the OLS method (ANOVA)	Bias in sum of squares
Y_{111}	35.675	35.675	1.86
Y_{123}	36.971	36.971	0.311
Y_{212}	32.376	32.376	12.22
Y_{233}	39.737	39.737	2.43
Y_{321}	40.425	40.425	11.61
Y_{332}	38.761	38.761	9.818

Here, six different positions, *viz*., Y_{111}, Y_{123}, Y_{212}, Y_{233}, Y_{321} and Y_{332} were considered as single missing and each position was estimated through ANCOVA model as well as ANOVA model. However, the estimated values using these methods are same. It is interesting to observe that the bias presents in sum of square of treatment for different positions are varying widely.

Table 3.1.2 shows the result of estimated missing values for six different positions (Y_{112}, Y_{131}, Y_{223}, Y_{232}, Y_{313} and Y_{321}) as single missing observation for the year 2014-15.

Table 3.1.2: Estimated value and bias in SS of the treatment for 2014-15:

Position	Estimated value by ANCOVA	Estimated value by the OLS method (ANOVA)	Bias in sum of squares
Y_{112}	34.428	34.428	1.628
Y_{131}	37.688	37.688	6.27
Y_{223}	37.025	37.025	1.493
Y_{232}	36.631	36.631	1.33
Y_{313}	32.312	32.312	0.459
Y_{321}	35.573	35.573	0.043

The bias in some of square of treatment is also calculated as given in table 3.1.1.

The results of table 3.1.1 and 3.1.2 reveal that the estimated values of missing observations for two different methods are same. The precision of the methods has been judged by comparing the values of MSE of the above mentioned two analysis procedures. The results of such comparisons for the year 2013-14 are shown in table 3.1.3.

Table 3.1.3: Mean square error of two methods for 2013-14

Position	MSE (ANCOVA)	MSE (OLS method)	Precision % of ANCOVA model
Y_{111}	5.72	5.846	2.17
Y_{123}	5.587	5.61	0.37
Y_{212}	5.646	6.46	14.43
Y_{233}	4.79	4.96	3.37
Y_{321}	5.37	6.145	14.41
Y_{332}	5.74	6.398	11.39

From the table 3.1.3, it is observed that for every position MSE is lower for ANCOVA method compared to the method of OLS. It has been observed that up to 14.43% MSE can be reduced by using the ANCOVA model for the position Y_{212} of the experiment under consideration. Table 3.1.4 also shows the gain in precision percentages of ANCOVA model over ANOVA model for the year 2014-15.

Table 3.1.4: Mean square error of two methods for 2014-15:

Position	MSE (ANCOVA)	MSE (OLS method)	Precision % of ANCOVA model
Y_{112}	1.30	1.41	8.32
Y_{131}	1.32	1.74	31.74
Y_{223}	1.16	1.261	8.57
Y_{232}	1.357	1.446	6.54
Y_{313}	1.015	1.045	3.02
Y_{321}	1.015	1.018	0.23

Here also we can observe that MSE for least-square method is higher for every position under consideration than ANCOVA method. In present experiment, the MSE of ANCOVA for the position Y_{131} has highest precision (31.74%) over the ordinary least square by ANOVA method.

3.2. Estimation of two missing observations

The results from the experiment mentioned in section 2.4 are now analysed to compare the two methods i.e., ANCOVA model and ANOVA model for estimation of two missing observations.

Table 3.2.1: Estimated values and bias in sum of squares for two missing observations in 2013-14

Position	Estimated value by ANCOVA	Estimated value by the OLS method (ANOVA)	Bias in sum of squares
Y_{123} & Y_{311}	37.26 & 536.333	37.263 & 36.331	0.080
Y_{132} & Y_{233}	38.034 & 39.664	38.031 & 39.661	5.237
Y_{121} & Y_{132}	37.742 & 38.144	37.740 & 38.141	4.382
Y_{122} & Y_{312}	35.688 & 36.853	35.693 & 36.858	0.950
Y_{213} & Y_{333}	35.520 & 40.857	35.526 & 40.861	14.254

We consider five different positions under two missing observations situation, i.e., Y_{123} & Y_{311} (random pair); Y_{132} & Y_{233} (pair of same 2nd factor); Y_{121} & Y_{132} (pair of same 1st factor); Y_{122} & Y_{312} and Y_{213} & Y_{333} (pairs of same block). We observed that the estimated values by both the methods are more or less same; there are no such significant differences. The amount of bias are also shown in the table and for the pair (Y_{213} & Y_{333}) positions, the bias is maximum, i.e., 14.254.

Table 3.2.2: Estimated values and bias in sum of squares for two missing observations in 2014-15

Position	Estimated value by ANCOVA	Estimated value by the OLS method (ANOVA)	Bias in sum of squares
Y_{123} & Y_{311}	36.162 & 33.445	36.160 & 33.443	3.623
Y_{132} & Y_{233}	37.953 & 35.370	37.951 & 35.368	6.341
Y_{121} & Y_{132}	35.781 & 37.797	35.778 & 37.795	6.047
Y_{122} & Y_{312}	36.365 & 33.114	36.369 & 33.119	6.142
Y_{213} & Y_{333}	34.406 & 37.114	34.412 & 37.119	3.703

For 2014-15, the positions are same as that of the previous year. Here also the same results have been observed. The highest amount of bias has been observed in case of the pair (Y_{132} & Y_{233}) position which is 6.341.

We mainly compare the methods on the basis of MSE which have been discussed below.

Table 3.2.3: Mean square error of two methods for two missing observations for 2013-14

Position	MSE (ANCOVA)	MSE (OLS method)	Precision % of ANCOVA model
Y_{123} & Y_{311}	5.293	5.301	0.1511
Y_{132} & Y_{233}	3.359	3.367	0.2590
Y_{121} & Y_{132}	3.946	3.954	0.1977
Y_{122} & Y_{312}	5.189	5.175	-0.2583
Y_{213} & Y_{333}	5.188	5.173	-0.2457

From the table 3.2.3, we observed that for the first three positions, *i.e.*, random pair, pair from the same 2nd factor and the pair from the same 1st factor the error mean square is less for ANCOVA than ANOVA. But the positions from the same block have higher error mean square for ANCOVA than ANOVA.

Table 3.2.4: Mean square error of two methods for two missing observations for 2014-15

Position	MSE (ANCOVA)	MSE (OLS method)	Precision % of ANCOVA model
Y_{123} & Y_{311}	1.422	1.429	0.4852
Y_{132} & Y_{233}	1.090	1.097	0.7067
Y_{121} & Y_{132}	1.434	1.441	0.5371
Y_{122} & Y_{312}	1.230	1.218	-0.9998
Y_{213} & Y_{333}	1.316	1.303	-0.9954

For the year 2014-15, the same pattern of results has been observed.

4. CONCLUSION

The above discussions have lead to a number of conclusions. Firstly, ANCOVA method can be used efficiently to estimate the missing values in factorial experiments. Secondly, for single missing observation, both the ANCOVA method and the ANOVA method estimate the same value for the missing observation but the ANCOVA method has more precision as the error mean square is less than the ANOVA method and there is no such bias in treatment sum of squares as that of the ANOVA method. So we can conclude that, if the data set has single missing value problem then one should analyze the data through ANCOVA model. Lastly, for two missing observations, both ANCOVA and ANOVA methods estimate the same value for every position and ANCOVA model is more precise than the ANOVA model as MSE is less for the first method than the second. But when two observations are missing from the same block, then the ANOVA model is more precise and should be used to estimate the missing values.

REFERENCES

Allan, F. E. and Wishart, J.1930. A Method of Estimating the Yield of a Missing Plot in Field Experimental Work. *Journal of Agricultural Science*, **20(3)**:390-406.

Anderson, R. L.1946. Missing-Plot Techniques. *Biometrics Bulletin*, **2(3)**:41-47.

Bartlett, M. S.1937. Some Examples of Statistical Methods of Research in Agriculture and Applied Biology. *Journal of the Royal Statistical Society*, **4(2)**:137-83

Coons, I.1957. The Analysis of Covariance as a Missing Plot Technique. *Biometrics*, **13(3)**:387-405.

Yates, F.1933. The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture*, **1(2)**:129-42.