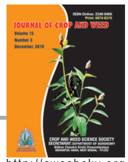
Journal of Crop and Weed, 16(3): 198-202 (2020)

ISSN- O : 2349 9400 ; P : 0974 6315



http://cwssbckv.org
www.cropandweed.com

# Application of ANCOVA Model for estimation of missing observation(s) and adjusted treatment means in split plot layout

**M. SAHA, J. S. BASAK AND A. MAJUMDER**

*Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya*
*Mohanpur- 741252, Nadia, West Bengal*

***Received : 05.12.2020 ; Revised : 30.08.2020 ; Accepted : 05.10.2020***

**DOI :** https://doi.org/10.22271/09746315.2020.v16.i3.13687

## ABSTRACT

*In agriculture, split plot experiments are becoming very popular among the researchers to quantify the responses of many factors through a single experiment. But sometimes, data loss in an experiment leads to major problem for analysis of experiment. Proper estimation of these missing observations from the data set leads to restore the orthogonal property of the designed experiments. The present study aims to apply the ANCOVA model in split plot layout with missing observation(s) as well as to compare different treatment mean values on the basis of adjusted mean values using regression coefficients of concomitant variables on response variables.The application of the above model for estimation of missing values (single and two) and analysis methods are showing very optimistic results in the field experiments on four yield parameters of maize during 2017-18. The absolute difference between the estimated missing value and the actual value is negligible in most of the cases. Moreover, the analysis by ANCOVA model with one and two missing observations shows that the adjusted treatment means along with their first order interactions (Federer, 1975) are same with the original data set which contains no such missing observations.*

***Keywords***: Adjusted treatment mean, ANCOVA, missing observations and split plot technique

Sometimes missing observations or presence of outliers create major information loss of any scientific experiment which leads to inaccurate variances. The analysis of missing plot technique follows some popular methods like principles of least square, method of fitting constants, analysis of covariance or ANCOVA, method of iteration etc. Among these, ANCOVA is a widely accepted method (Coons, 1957) because firstly, some unexplained variance can be explained in terms of covariates and within-group error variance can be reduced. Secondly, ANCOVA is ideally suited to remove the bias of any variable which influence the measures of dependent variable. The main objectives of this study are to apply the ANCOVA model on the field experiments in split plot lay out with single as well as twomissing observations and to compare different treatment mean values (main plot, sub plot and Interactions) on the basis of adjusted mean values using regression coefficients of concomitant variables on response variables.

Cochran (1957) described the application and the standard guidelines for ANCOVA in a special issue of Biometrics. In the same issue, Coons (1957) discussed the application of ANCOVA model for estimation of missing observations in multifactor experiments. Almimi *et al*. (2008) reported that inserting estimates for the missing observations from split-plot designs restore their balanced or orthogonal structure. Federer (1975, 1977, 1992) in different studies discussed the computational procedure of adjusted treatment means' main effect and interactions.

## MATERIALS AND METHODS

The crop maize (Variety: Shrestha) was cultivated in the Instructional farm, Jaguli, Bidhan Chandra Krishi Viswavidyalaya for the years 2017-18. A split plot design with three factors : (D) Date of sowing, (S) Spacing and (N) Dose of nitrogen was performed for this study. Where the different dates of sowing ($D_1$, $D_2$, $D_3$) is used as the whole plot factor and the factorial combination of spacing ($S_1$, $S_2$, $S_3$) with different doses of nitrogen ($N_1$, $N_2$, $N_3$) was plotted as sub plot factor. Green cob yield data at harvest were recorded as dependent variable. Single and two missing observations, both the cases were considered and analysis was performed following Coons (1957).

### *For single missing observation*

Determining the dependent variable Y, insert the value of zero in all of the cell where the observation is missing. For one missing value we have constructed only one concomitant variable X such as : $X = 0$ iff $Y \neq 0$, $X = -n$ iff $Y = 0$

Here -n is choosen because it easily makes the sum of squares of the concomitant variable to n×degrees of freedom. The ANCOVA table for one missing observation consisting of three factors are given in table 1. The levels of D, S and N are p, q and s,

*Email: mriganka.19@gmail.com*

respectively and the number of replication is r. The computational procedure of analysis of the covariance automatically provides unbiased tests of significance.The regression coefficients for main plot and sub plot factors are estimated by the following formulae :

$$\hat{\beta}w = E(I)_{yx} / E(I)_{xx} \quad \text{and} \quad \hat{\beta}s = E(II)_{yx} / E(II)_{xx}$$

Now to estimate the missing value $\hat{y}$ , multiply this estimate of the regression coefficient by n :

$$\hat{y}=n\hat{\beta}_s$$

The following formulae of Federer (1975) are used to calculate the adjusted means :

Adjusted treatment means for main plot treatment :

(i) For factor D : $\bar{y}'_{i...} = \bar{y}'_{i...} - \hat{\beta}_w (\bar{x}_{i...} - \bar{x}_{i...})$

Adjusted treatment means for sub plot treatments:

(ii) For factor S : $\bar{y}'_{j..} = \bar{y}'_{j..} - \hat{\beta}_s (\bar{x}_{j..} - \bar{x}_{...})$

(iii) For factor N : $\bar{y}'_{k..} = \bar{y}'_{k..} - \hat{\beta}_s (\bar{x}_{k..} - \bar{x}_{...})$

First order interactions of three main treatments (D, S and N):

(iv) adj.$\bar{y}_{ij...} = \bar{y}_{ij...} = \bar{y}_{ij} - \hat{\beta}_s (\bar{x}_{ij...} - \bar{x}_{i...}) - \hat{\beta}_w (\bar{x}_{i...} - \bar{x}_{...})$

(v) adj. $\bar{y}_{i.k} = \bar{y}'_{i.k} = \bar{y}_{i.k} - \hat{\beta}_s (\bar{x}_{i.k} - \bar{x}_{i...}) - \hat{\beta}_w (\bar{x}_{i...} - \bar{x}_{...})$

(vi) adj. $\bar{y}_{j.k.} = \bar{y}'_{.jk.} = \bar{y}_{.jk.} - \hat{\beta}_s (\bar{x}_{.jk.} - \bar{x}_{.j...}) \hat{\beta}_s (\bar{x}_{.j...} - \bar{x}_{...})$

or, $\bar{y}'_{jk.} = \bar{y}_{.jk.} - \hat{\beta}_s (\bar{x}_{.jk.} - \bar{x}_{...})$

**Table 1: ANCOVA table for one missing observation**

| S.O.V. | d.f. | Sum of products | | | Adj. d.f. | Adj. Sum of squares (adj. YY) | Mean Square |
|---|---|---|---|---|---|---|---|
| | | YY | YX | XX | | | |
| Replication(R) | (r-1) | $R_{yy}$ | $R_{yx}$ | $R_{xx}$ | | | |
| Factor D | (p-1) | $D_{yy}$ | $D_{yx}$ | $DA_{xx}$ | (p-1) | $D'_{yy}$ | (adj.YY/ adj. d.f.) |
| Error (I) | (r-1)(p-1) | $E(I)_{yy}$ | $E(I)_{yx}$ | $E(I)_{xx}$ | (r-1)(p-1) | E(I)′yy | |
| Factor S | (q-1) | $S_{yy}$ | $S_{yx}$ | $S_{xx}$ | (q-1) | S′yy | |
| D × S | (p-1)(q-1) | $DS_{yy}$ | $DS_{yx}$ | $DS_{xx}$ | (p-1)(q-1) | DS′yy | |
| Factor N | (s-1) | $N_{yy}$ | $N_{yx}$ | $N_{xx}$ | (s-1) | N′yy | |
| D × N | (p-1)(s-1) | $DN_{yy}$ | $DN_{yx}$ | $DN_{xx}$ | (p-1)(s-1) | DN′yy | |
| S × N | (q-1)(s-1) | $SN_{yy}$ | $SN_{yx}$ | $SN_{xx}$ | (q-1)(s-1) | SN′yy | |
| D × S × N | (p-1)(q-1)(s-1) | $DSN_{yy}$ | $DSN_{yx}$ | $DSN_{xx}$ | (p-1)(q-1)(s-1) | DSN′yy | |
| Error (II) | p(r-1)(qs-1) | $E(II)_{yy}$ | $E(II)_{yx}$ | $E(II)_{xx}$ | p(r-1)(qs-1)-1 | E(II)′yy | |
| Total | pqrs-1 | | | | | | |

***For two missing observations***

At first, consider the original data as the dependent variable Y and insert the value of zero in the cellswhere the observations are missing. Then define a concomitant variable $X_m$ for each missing observation, where:

$X_m = 0$ iff $Y \neq 0$ and $X_m = -n$ iff $Y = 0$ for m = 1,2.

For the conminant variable -n is chosen because it simply leads to the consequence that the sum of squares for the concomitant variable is equals to n × degrees of freedom. The computations required to obtain the sum of products $\Sigma X_m X_n$ and $\Sigma X_m Y$, since each $X_m$ is associated with a single missing value and therefore has only one non-zero cell. Now to get the estimates of the regression coefficients for main plots $(\hat{\beta}_{1w}, \hat{\beta}_{2w})$ and sub plots $(\hat{\beta}_{1s}, \hat{\beta}_{2s})$, solve these four equations given below :

$$E(I)_{X_1X_1}\hat{\beta}_{1w}+E(I)_{X_1X_2}\hat{\beta}_{2w}= E(I)_{X_1Y}$$

$$E(I)_{X_2X_1}\hat{\beta}_{1w}+ E(I)_{X_2X_2}\hat{\beta}_{2w}= E(I)_{X_2Y}$$

$$E(II)_{X_1X_1}\hat{\beta}_{1s}+ E(II)_{X_1X_2}\hat{\beta}_{2s}= E(II)_{X_1Y}$$

$$E(II)_{X_2X_1}\hat{\beta}_{1s}+E(II)_{X_2X_2}\hat{\beta}_{2s}= E(II)_{X_2Y}$$

After getting the regression coefficients each missing observation $Y_i$ is estimated by :

$\hat{Y}_m = n \times \hat{\beta}_{ms}$ for m = 1 and 2.

Using the estimated regression coefficients, obtained in the terms of $\hat{\beta}_{1w}$, $\hat{\beta}_{2w}$, $\hat{\beta}_{1s}$ and $\hat{\beta}_{2s}$, adjusted values of all treatment means based on their individual performances and their first order interactions can be calculated.

## RESULTS AND DISCUSSION

Here, in table-2, the results are given for the experiment on green cob yield at harvest in 2017-18 for single missing observation and two missing observations. The observation of 2nd date of sowing ($D_2$) in main plot in combination of 3rd type of spacing with 1st dose of nitrogen ($S_3N_1$) in sub plot in the 3rd replication ($R_3$) is assumed as single missing observation. The absolute difference is only 0.25, which is nearly similar to the original data. Absolute difference is calculated by the absolute value of difference between estimated and original observation.The ratio of error mean square between ANOVA and ANCOVA are 1.09 and 0.98 in case of error(I) and error(II) respectively. Where the ratio are too close to 1.00, indicate there are more or less similar error in ANOVA and ANCOVA.But the ratio more than 1.00 indicates error in ANOVA is more than error in ANCOVA. For two missing observations, 2nd date of sowing ($D_2$) in main plot in combination of 1st type of spacing with 2nd dose of nitrogen ($S_1N_2$) in sub plot in the 3rd replication ($R_3$) and 3rd date of sowing ($D_3$) in main plot in combination of 3rd type of spacing with 3rd dose of nitrogen ($S_3N_3$) in sub plot in the 1st replication ($R_1$) are assumed to be lost. The absolute differences for the first and second missing observation were only 0.31 and 0.81 respectively. It indicates that estimation is more or less perfect for both the missing observations. The ratio of error mean square between ANOVA and ANCOVA are 4.46 and 0.95 in case of error (I) and error (II), respectively. So it is clear that analysis through ANCOVA model for two missing observations has shown excellent performance for error (I) and more or less similar result for error (II) over ANOVA models of the split plot design.

**Table 2: Single and two missing observations' estimation on green cob yield at harvest in 2017-18**

| No. of missing obs. | Estimated missing obs. | Original data | Absolute difference | EMS(I) from ANOVA / EMS (I) from ANCOVA | EMS(II) from ANOVA / EMS (II) from ANCOVA |
|---|---|---|---|---|---|
| Single missing | 33.25 | 33 | 0.25 | 1.09 | 0.98 |
| Two missing | 24.69 | 25 | 0.31 | 4.46 | 0.95 |
| | 38.19 | 39 | 0.81 | | |

An attempt has also been made to see the overall performance of the treatments for experiment without missing observations and experiments with missing observations. Table-3 present the adjusted mean values of three main factors *viz*., Date of sowing (D), Spacing (S) and Doses of nitrogen (N) of experiments and actual or original mean values on green cob yield. The results show that the actual positions of the treatments (date of sowing, spacing and dose of nitrogen) for both the columns of missing and original are same. The results for single missing observation and two missing observations on green cob yield are also shown in fig.1 and 2 respectively, for visual understanding. Next, the overall performances of the first order interactions of three main treatments (D, S and N) for the experiment without missing observations and experiments with missing observations are presented in table 4. The results also show that the actual positions of the first order interaction of treatments (date of sowing, spacing and dose of nitrogen) for both the columns of missing and original are same. The results of the table for single missing observation and two missing observations are also shown in fig. 3 and 4, respectively for visual understanding. Thus, it can be concluded that the present methodology for analysis with missing observation does not make any unjustified result for individual effect as well as interactions also.

**Table 3: Individual effect of all treatments on green cob yield at harvest in'17-18 (t ha$^{-1}$)**

| Treatment | Adj. mean from ANCOVA (single missing) | Adjusted mean from ANCOVA (two missing) | Original mean from ANOVA |
|---|---|---|---|
| $D_1$ | 26.262 | 25.956 | 26.704 |
| $D_2$ | 29.401 | 29.170 | 29.741 |
| $D_3$ | 25.929 | 25.319 | 26.370 |
| $S_1$ | 21.330 | 20.953 | 21.741 |
| $S_2$ | 24.701 | 24.335 | 25.111 |
| $S_3$ | 35.562 | 35.157 | 35.963 |
| $N_1$ | 25.117 | 24.742 | 25.519 |
| $N_2$ | 26.700 | 26.323 | 27.111 |
| $N_3$ | 29.774 | 29.379 | 30.185 |

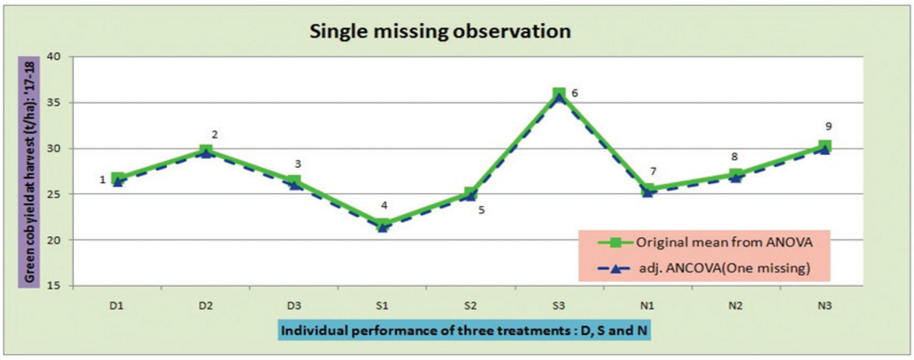



**Fig. 1: Adjusted mean values of three main treatments individually for green cob yield at harvest in 2017-18 in split plot layout for single missing observation**

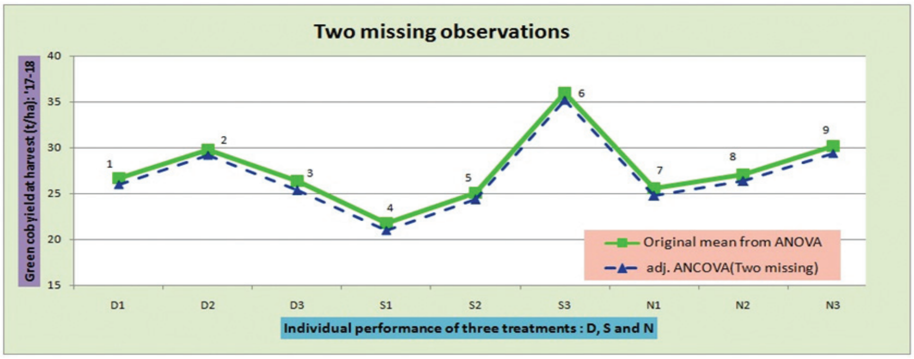



**Fig. 2: Adjusted mean values of three main treatments individually for green cob yield at harvest in 2017-18 in split plot layout for two missing observations**

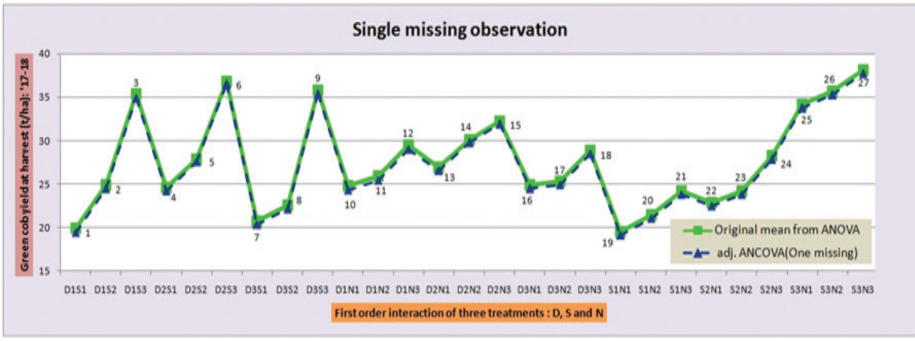



**Fig. 3: Adjusted mean values of first order interactions of all the treatments for green cob yield at harvest in 2017-18 in split plot layout for single missing observation**

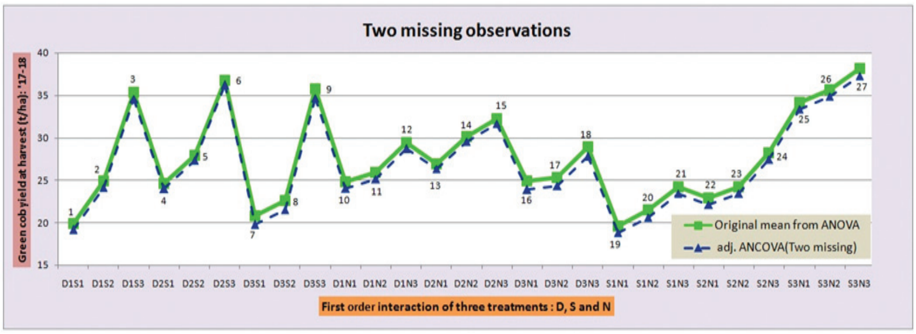



**Fig. 4: Adjusted mean values of first order interactions of all the treatments for green cob yield at harvest in 2017-18 in split plot layout for two missing observations**

**Table 4: First order interactions of all the treatments on green cob yield at harvest in 2017-18 (t ha$^{-1}$)**

| Treatment interaction | Treatment combination | Adj. mean from ANCOVA (single missing) | Adj. mean from ANCOVA (Two missing) | Original mean from ANOVA |
|---|---|---|---|---|
| D × S | $D_1S_1$ | 19.448 | 19.141 | 19.889 |
| | $D_1S_2$ | 24.448 | 24.141 | 24.889 |
| | $D_1S_3$ | 34.892 | 34.585 | 35.334 |
| | $D_2S_1$ | 24.207 | 23.962 | 24.556 |
| | $D_2S_2$ | 27.540 | 27.330 | 27.889 |
| | $D_2S_3$ | 36.457 | 36.219 | 36.778 |
| | $D_3S_1$ | 20.336 | 19.756 | 20.778 |
| | $D_3S_2$ | 22.114 | 21.534 | 22.556 |
| | $D_3S_3$ | 35.336 | 34.666 | 35.778 |
| D × N | $D_1N_1$ | 24.336 | 24.030 | 24.778 |
| | $D_1N_2$ | 25.448 | 25.141 | 25.889 |
| | $D_1N_3$ | 29.003 | 28.696 | 29.445 |
| | $D_2N_1$ | 26.568 | 26.330 | 26.889 |
| | $D_2N_2$ | 29.762 | 29.518 | 30.111 |
| | $D_2N_3$ | 31.873 | 31.663 | 32.222 |
| | $D_3N_1$ | 24.448 | 23.867 | 24.889 |
| | $D_3N_2$ | 24.892 | 24.3116 | 25.334 |
| | $D_3N_3$ | 28.448 | 27.777 | 28.889 |
| S × N | $S_1N_1$ | 19.145 | 18.779 | 19.556 |
| | $S_1N_2$ | 21.034 | 20.633 | 21.445 |
| | $S_1N_3$ | 23.812 | 23.446 | 24.222 |
| | $S_2N_1$ | 22.478 | 22.113 | 22.889 |
| | $S_2N_2$ | 23.812 | 23.446 | 24.222 |
| | $S_2N_3$ | 27.812 | 27.446 | 28.222 |
| | $S_3N_1$ | 33.728 | 33.335 | 34.111 |
| | $S_3N_2$ | 35.256 | 34.890 | 35.667 |
| | $S_3N_3$ | 37.701 | 37.245 | 38.111 |

In most of the cases of estimation of missing observation(s), the absolute difference between the original and estimated values are very negligible. It is also observed that the ANCOVA model is very much useful and relevant to the present study. Error (II) term also reduces in case of ANCOVA model than ANOVA. The adjusted mean values of the effects of different main factors and interactions for missing observations by ANCOVA model are also showing similar trend with that of the original mean values for no missing observation.

## REFERENCES


Ahmed, L.A. 2016. Missing values estimation comparison in split-plot design. *Int. J.Comput. Inf. Technol.*, **5**: 337-344.

Allan, F.E. and Wishart, J. 1930. A method of estimating yield of missing plot in field experimental work. *J.Agril. Sci.*,**20**(3): 399-406.

Anderson, R.L. 1946. Missing plot techniques. *J.Biom. American Stat. Asso.*, **2**(3): 41-47.

Clason, D.L. and Mundfrom, D.J. 2012. Adjusted means in analysis of covariance: Are they meaningful? *Multiple Linear Regression Viewpoints*, **38**: 8-15.

Cochran, W.G. 1957. Analysis of covariance: Its nature and uses. *Biom.*,**13**(3): 261-281.

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

Dutta, A., Majumder, A., Nishad, D. and Mandi, K. 2019. Estimation of missing observations of agricultural experiments under split plot set up. *J. Crop and Weed*, **15**(1): 102-107.

Federer, W.T. 1975. The misunderstood split plot. Applied Statistics (Ed.Gupta, R.P.) North-Holland Publishing Co., Amsterdam, Oxford, pp. 39.

Federer, W.T. 1977. Sampling, blocking and modelconsiderations for splitplot and split block designs. *Biom. J.*, **19**: 181-200.

Federer, W.T. and Meredith, M.P. 1992. Covariance analysis for split-plot and split-block designs.*The American Statistician*, **46**(2): 155-162.

Smith, H.F.1957. Interpretation of adjusted treatment means and regressions in analysis of ovariance. *Biom.*, **13**(3): 282-308.