Journal of Crop and Weed, 15(1): 102-107 (2019)

Estimation of missing observations of agricultural experiments under split plot set up

A. DUTTA, A. MAJUMDER, D. NISHAD AND [1]S.K. MANDI

Department of Agricultural Statistics, [1]Department of Agronomy
Faculty of Agriculture, Bidhan Chandra Krishi Viswavidyalaya, Mohanpur-741252, Nadia, West Bengal

Received : 03-08-2018 ; Revised : 12-01-2019 ; Accepted : 02-01-2019

ABSTRACT

Estimation of single, double and triple missing observations are done in agricultural field trials on effect of varieties of Lentil under Relay Cropping with Medium and Long Duration Rice Varieties in split plot set up at District Seed Farm, Bidhan Chandra Krishi Viswavidyalaya, Kalyani, West Bengal for consecutive two years (2014-15 and 2015-16). The missing values are estimated by three methods, viz., Coons (1957), Rubin (1972) and Haseman and Gaylor (1973). Three criteria viz., Absolute Error (AE), Mean Square Error (MSE) and Akaike Information Criterion (AIC) are used to judge the above three methods used for estimation of missing observations. The precision of three methods for single missing observation are same for both the years under study. But, it is observed that precision of Coons and Haseman and Gaylor are equally effective and both are better than Rubin's method for two and three missing observations. The results are similar for both the years under study.

Keywords **:** *Analysis of Covariance (ANCOVA) model; non- iterative technique to estimate missing observations*

Sometimes in agricultural field experiments, the observations get lost due to unavoidable circumstances or so much affected by some extraneous causes that it wouldn't be desirable to regard those observations as normal experimental observations. Such experimental data with missing observation (or observations) are generally analysed through the technique of missing plot. There are several methods of estimating the missing observation (or observations) like, minimising the error sum of squares, method of iteration, method of fitting of constants and analysis of the data with missing observations by the technique of analysis of covariance (ANCOVA) model. But the application of the methods including ANCOVA model for missing data analysis in multifactor experiments, specially, in split plot set up 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 squares and in addition he also obtained the correct least squares estimates of all estimable parameters. Actually, 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). Rubin (1972) reported a non- iterative algorithm for estimation of missing values in any experimental designs under ANOVA models. Haseman and Gaylor (1973) also reported a simpler non-iterative approach for obtaining missing observation estimates by solving a set of simultaneous linear equations. The method was derived for the two-way crossed classification and results for the P-way (P e" 2) crossed classification were also given by them.

Recently, Ahmed (2016) focused on a comparative discussion on missing yields in split plot set up with three methods, *viz*., Coons (1957), Rubin (1972) and Haseman and Gaylor (1973). The methods were judged by three well known measures, *viz*., absolute error (AE), mean square error (MSE) and Akaike Information Criterion (AIC) on a field trial in split plot set up. Despite ample availability of methods for estimation and analysis of missing observations in multifactor experiments, experimenters are not using the appropriate methods for the purpose.

The main objective of the present study is to compare the available methods for analysis of a split plot experiment when one or many observations are missing. In all, three available methods, *viz*., ANCOVA model (Coons, 1957), and two non- iterative methods like Rubin (1972) and Haseman and Gaylor (1973) are applied to field level experiments on lentil crop yield values for consecutive two years (2014-15 and 2015-16) under a split plot design set up for estimation of single, two and three missing observations. The above mentioned three methods *viz*., Coons (1957), Rubin (1972) and Haseman and Gaylor (1973) are compared among themselves for one, two, three missing observations.

Email: ankita.manoj1993@gmail.com

MATERIALS AND METHODS

Estimation of missing observations in Split-plot design with ANCOVA (Coons, 1957)

Firstly, ANCOVA model can be used to estimate the missing values in experimental designs. Coons (1957) gave analysis of covariance model to analyze the experiments with missing values. The technique employs the computational procedures of a multiple covariance analysis using one or more concomitant variables, X_i (i = 1, 2, 3, ..., m) as follows, where Y is considered as original observation:

When there are m missing values:

I. Put Y = 0 for all missing values.

II. Define m new variables denoting X_i (i = 1, 2, 3, ..., m) where: $X_i = 0$ iff $Y \neq 0$ and $X_i = -n$ iff Y = 0 for all i.

III. With more than one missing observations, a multiple covariance analysis is required.

The computations required to obtain the sum of products $\Sigma X_i X_j$ and $\Sigma X_i Y$, since each X_i is associated with a single missing value and therefore has only one non-zero cell. In computing $\Sigma X_i X_j$, two situations may be encountered.

a) The two missing values associated with X_i and X_j occur in the same level of the given source of variation. $\Sigma X_i X_j$ = n (Degree of freedom for the given source of variation).

b) The two missing values occur in the different levels of the given source of variation. $\Sigma X_i X_j$ = -n (Degree of freedom for the given source of variation)

IV. Compute the estimates of the regression coefficients $\left(\hat{\beta}_{12E}, \hat{\beta}_{22E}, \ldots\ldots, \hat{\beta}_{m2E}\right)$ by solving m equations:

$$E_{X_1X_1}\hat{\beta}_{12E} + E_{X_1X_2}\hat{\beta}_{22E} + \ldots\ldots + E_{X_1X_m}\hat{\beta}_{m2E} = E_{X_1Y}$$
$$E_{X_mX_1}\hat{\beta}_{12E} + E_{X_mX_2}\hat{\beta}_{22E} + \ldots\ldots + E_{X_mX_m}\hat{\beta}_{m2E} = E_{X_mY} \quad \ldots\ldots (1)$$

We estimate the missing values by the following formula : $\hat{Y}_i = n\hat{\beta}_{i2E}$, i = 1, 2, 3,..., m.

e.g. If there is only one missing value, then we introduce only one concomitant variable, X_1 and there will be a single equation i.e. $E_{X_1X_1}\hat{\beta}_{12E} = E_{X_1Y}$. The estimate of the missing value will $\hat{Y} = n\hat{\beta}_{12E}$.

Estimation of missing observation by Rubin's Method

In (1972) Rubin used non- iterative technique to estimate missing values and in a way that using least squares and make the sum of squares error equal to zero. The vector of estimated missing values, $X = -PR^{-1}$, where: P and X be the Vectors of order (1 × m).

The elements of P vector, i.e., $e_{ijk} = Y_{ijk} - \frac{Y_{ij.}}{b} - \frac{Y_{.jk}}{r} + \frac{Y_{.j.}}{br}$, where, Y_{ijk} be the missing value = 0; Y_{ij} be the total for main-plot containing the missing value in the block; $Y_{.jk}$ be the total of all sub units that receive the treatment combination (a_ib_j) which has the missing value(s); $Y_{.j}$ be the total of all observations that receive the i^{th} level of A; R be the Non-singular matrix (m × m). The diagonal elements of matrix R, will be, $r_{kk} = 1 - \frac{1}{b} - \frac{1}{r} + \frac{1}{br}$, and the off-diagonal elements, $r_{kk'} = \frac{1}{br}; (k \neq k')$.

Estimation of missing observation by Haseman and Gaylor method

Haseman and Gaylor (1973) suggested a non-iterative technique to estimate **m** missing values by solving **m** of simulations linear equations, the formula as follows:

$$(r-1)(b-1)Y_h + \sum_{g\neq h}^{m} Y_g[\psi_{gh}(A_3) - r\psi_{gh}(A_1) - b\psi_{gh}(A_2) = rT_h(A_1) + bT_h(A_2) - T_h(A_3) \ldots\ldots (2)$$

where, Y_h and Y_g are the missing values (g, h = 1, 2,…, m); r be the number of replicates and b is the number of levels of the sub-plot factor.

$\phi_{gh}(A_1)$ = 1, If Y_h and Y_g are of different levels of factor B, but from the same levels of factor A and in the same block.

0, Otherwise

$\phi_{gh}(A_2)$ =1, If Y_h and Y_g are of a particular level for the factors A and B.

0, Otherwise

$\phi_{gh}(A_3)$ = 1, If Y_h and Y_g are of the same level for Factor A.

0, Otherwise

$T_h(A_1)$ = Total for main unit containing the missing value

$T_h(A_2)$ = Total of all sub units that receive the treatment combination (a_ib_j).

$T_h(A_3)$ = Total of all observations that receive the i^{th} level of A.

Statistical measures for comparison of the methods

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 some statistical measurements are calculated to compare the methods for split plot design: absolute error (AE), mean square error (MSE), Akaike Information Criterion (AIC) etc.

Absolute error is the absolute of the difference between estimated missing value and real value, and calculated as follows:

$AE = |y_i - \hat{y}_i|$ where, y_i is real and $\hat{y}_i$ is estimated value.

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

Akaike information criterion (AIC) is a measure of the relative quality of statistical methods for a given set of data, is calculated as follows:

$$AIC = n \ln \sigma^2 + 2(k + 1)$$

where, σ^2 is Mean square of error, k is the number of variables in the model and n is total number of observation.

Experimental details

The split plot experiment was conducted by Mr. Sanjib Kumar Mandi for his Master's dissertation work under Dr. R. Nath, Professor of Agronomy Department, BCKV at District Seed Farm, Bidhan Chandra Krishi Viswavidyalaya, AB Block, Kalyani on "Effect of Varieties and Seed Rates on Growth, Yield and Yield Attributes of Lentil under Relay Cropping with Medium and Long Duration Rice Varieties". The geographical location of the farm is 22°93' N latitude, 88°53' E longitude and 9.75 m above mean sea level for consecutive two years (2014-15 and 2015-16). The experiments were conducted on a medium land, well-drained Gangetic alluvial soil (order: Inceptitisol), which belonged to the class of clayey loam with medium fertility and almost neutral in reaction. The main-plot and sub-plot treatments under study are given as follows:

Table 1: The experiment details for a 3x4 Split-plot set-up with three replications

Main plot factor: variety of lentil (V)	Sub-plot factor: seed rate of lentil (S)
V1: PL 6	S1: Seed rate of 50 kg ha^{-1} as relay cropping of lentil
V2: WBL 77	S2: Seed rate of 60 kg ha^{-1} as relay cropping of lentil
V3: NDL 1	S3: Seed rate of 70 kg ha^{-1} as relay cropping of lentil
	S4: Seed rate of 80 kg ha^{-1} as relay cropping of lentil

The seed yield of lentil (kg ha^{-1}) was taken for our study from the above mentioned experiment.

RESULTS AND DISCUSSION

Three different situations are studied for estimation of missing observations in case of split plot design, i.e. a) single missing observation, b) two missing observations, and c) three missing observations.

Here three different methods are applied to estimate the values of missing observations for each situation. The results from the experiments described in section 2.5 are presented for two consecutive years (2014-15 and 2015-16).

Estimation of single missing observation

Three different methods *viz*., Coons (ANCOVA), Haseman and Gaylor's and Rubin's methods were used to estimate the single missing observations. The precision of the experiment has been judged by three criterion *viz*., absolute error (AE), mean square error (MSE) and Akaike Information Criterion (AIC) as mentioned above. Table 2 presents the estimated missing values and the respective criterion values for the year 2014-15.

Here the two different positions *viz*. Y_{121} and Y_{142} were considered as single missing observation in table 2. The results show that the estimated values by all the three methods are same and the values of the judgement criteria are also equal for the methods.

The table 3 also shows the similar results for two different positions (Y_{111} and Y_{223}) for the year 2015-16 and proves that the three methods are equally efficient for estimation of single missing observation.

Table 2: Estimated single missing value and criteria values for 2014-15

Position	Methods	Estimated Value	AE	MSE	AIC
*Y_{121}	Coons	1074.38	19.37	12906.50	350.75
	Haseman & Gaylor	1074.38	19.37	12906.50	350.75
	Rubin	1074.38	19.37	12906.50	350.75
Y_{142}	Coons	1048.75	17.50	12908.54	350.76
	Haseman & Gaylor	1048.75	17.50	12908.54	350.76
	Rubin	1048.75	17.50	12908.54	350.76

*Note: * Y_{ijk} indicates the yield of i^{th} main plot in j^{th} sub plot in k^{th} replication, i = 1, 2, 3; j = 1, 2, 3, 4 and k = 1, 2,3*

Table 3: Estimated single missing value and criteria values for 2015-16

Position	Methods	Estimated Value	AE	MSE	AIC
*Y_{111}	Coons	1016.72	10.85	4017.89	308.75
	Haseman & Gaylor	1016.72	10.85	4017.89	308.75
	Rubin	1016.72	10.85	4017.89	308.75
Y_{223}	Coons	1028.35	11.66	4017.30	308.74
	Haseman & Gaylor	1028.35	11.66	4017.30	308.74
	Rubin	1028.35	11.66	4017.30	308.74

*Note: * Y_{ijk} indicates the yield of i^{th} main plot in j^{th} sub plot in k^{th} replication, i = 1, 2, 3; j = 1, 2, 3, 4 and k = 1, 2,3*

Estimation of two missing observations

Table 4 presents the results for two missing observations in 2014-15. Here also we consider the pair of missing values from three different pair positions. Firstly, one pair from same main plot and same replication but in two different sub-plots (Y_{111} and Y_{141}) is considered. Secondly, the pair is selected from different main plots and different replications but in same sub-plot (Y_{113} and Y_{312}). Lastly, the third pair is selected from same main plot and sub-plot but different replications (Y_{111} and Y_{112}).

The missing pairs are estimated by the three methods mentioned earlier. From the table it has been observed that the values of absolute error for every position for Coons method and Haseman and Gaylor's method are lesser from that of the Rubin's method. The MSE and AIC values are also showing the same results, whereas, Rubin's method gives the higher values for every position compared to remaining two methods *viz*., Coons method and Haseman and Gaylor's method. It is noted that for the pair position Y_{111} and Y_{112}, the deviation is maximum. The results imply that Coons method and Haseman and Gaylor's method are equally efficient and both methods minimised the residual errors. But Rubin's Method does not show better result than the other two.

Table 4: Estimated two missing values and criteria values for 2014-15

Position	Methods	Estimated Value	AE	MSE	AIC
*Y_{111} Y_{141}	Coons	1123.441092.19	53.4464.06	13436.09	352.21
	Haseman & Gaylor	1123.441092.19	53.4464.06	13436.09	352.21
	Rubin	658.04608.04	411.96548.21	22835.45	371.30
Y_{113} Y_{312}	Coons	954.7921130.04	48.542150.16	12946.65	350.87
	Haseman & Gaylor	954.7921130.04	48.542150.16	12946.65	350.87
	Rubin	788.35998.65	117.9281.55	14351.85	354.58
Y_{111} Y_{112}	Coons	1087.5979.17	17.5114.58	13242.37	351.68
	Haseman & Gaylor	1087.5979.17	17.5114.58	13242.37	351.68
	Rubin	540.36345.36	529.64748.39	24313.61	373.56

*Note: * Y_{ijk} indicates the yield of i^{th} main plot in j^{th} sub plot in k^{th} replication, i = 1, 2, 3; j = 1, 2, 3, 4 and k = 1, 2,3*

Table- 5 presents also the results for two missing observations for the year 2015-16. Exactly, the similar types of results of table 4 are repeated in table 5. The pairs of missing observations are again considered for three different pair of positions in the experiment. The position combinations are selected as (y_{111} and y_{131}), (y_{121} and y_{322}) and (y_{111} and y_{112}) as done in table 4. Here also, the AE, MSE, AIC values are higher in case of Rubin's method. It has been observed that there is maximum deviation of AE, MSE and AIC for the pair position Y_{111} and Y_{112}. Considering the tables 4 and 5, it is observed that the pair position Y_{111} and Y_{112} provides the maximum value for Rubin's method. The results have shown that the methods described by Coons and Haseman and Gaylor are similar and both are more effective than the third method described by Rubin for all three criteria under study.

Table 5: Estimated two missing values and criteria values for 2015-16

Position	Methods	Estimated Value	AE	MSE	AIC
*Y_{111} Y_{131}	Coons	1034.71008.6	28.8353.97	4188.02	310.24
	Haseman & Gaylor	1034.71008.6	28.8353.97	4188.02	310.24
	Rubin	604.68562.94	401.19391.69	12181.6	348.67
Y_{121} Y_{322}	Coons	971.2781276.72	13.10235.78	4227.29	310.58
	Haseman & Gaylor	971.2781276.72	13.10235.78	4227.29	310.58
	Rubin	780.1631146.69	204.2294.25	5897.12	322.56
Y_{111} Y_{112}	Coons	1085.21099.6	79.33136.93	3829.39	307.02
	Haseman & Gaylor	1085.21099.6	79.33136.93	3829.39	307.02
	Rubin	455.19481.16	550.68481.51	16009.74	358.51

*Note: * Y_{ijk} indicates the yield of i^{th} main plot in j^{th} sub plot in k^{th} replication, i = 1, 2, 3; j = 1, 2, 3, 4 and k = 1, 2, 3*

Table 6: Estimated three missing values and criteria values for 2014-15

Position	Methods	Estimated Value	AE	MSE	AIC
*Y_{111}, $Y_{133}Y_{141}$	Coons	1188.35, 771.59, 1157.10	118.35, 259.66, 0.85	12271.68	348.94
	Haseman & Gaylor	1188.35, 771.59, 1157.10	118.35, 259.66, 0.85	12271.68	348.94
	Rubin	670.00, 947.50, 620.00	400.00, 83.75, 536.25	23625.63	372.52
Y_{111}, $Y_{213}Y_{312}$	Coons	1144.79, 1168.75, 1130.04	74.79, 237.50, 150.16	11821.63	347.59
	Haseman & Gaylor	1144.79, 1168.75, 1130.04	74.79, 237.50, 150.16	11821.63	347.59
	Rubin	857.21, 885.96, 839.51	202.79, 520.29, 440.69	20057.49	366.63
Y_{111}, $Y_{233}Y_{342}$	Coons	1144.79, 1166.67, 1098.44	74.79, 183.33, 151.56	12567.39	349.79
	Haseman & Gaylor	1144.79, 1166.67, 1098.44	74.79, 183.33, 151.56	12567.39	349.79
	Rubin	862.27, 888.52, 806.65	202.79, 520.29, 440.69	20645.32	367.67

*Note: * Y_{ijk} indicates the yield of i^{th} main plot in j^{th} sub plot in k^{th} replication, i = 1, 2, 3; j = 1, 2, 3, 4 and k = 1, 2,3*

Estimation of three missing observations

Lastly the study extended for three missing values from four different positional combinations for consecutive 2 years, 2014-15 and 2015-16, presented in table 6 and table 7, respectively.

Here also we consider three different combinations of three missing values. First combination is from the same main plot but different sub-plot and replication (Y_{111}, Y_{133} and Y_{141}). The second combination has only sub- plot in common (Y_{111}, Y_{213} and Y_{312}). The third combination is considered having no common factors or replication (Y_{111}, Y_{233} and Y_{342}). It is noted that for the first combination the MSE and AIC values are comparatively larger than other combinations in case of Rubin's method.

Table- 7 shows the similar results for the year 2015-16 for the same combinations of positions of three missing values.

Results of the tables- 6 and 7 also confirm that the methods given by Coons and Haseman and Gaylor are

Table 7: Estimated three missing values and criteria values for 2015-16

Position	Methods	Estimated Value	AE	MSE	AIC
*Y_{111}, $Y_{133}Y_{141}$	Coons	1002.47, 1063.63, 905.54	3.40, 0.75,43.13	4498.002	312.81
	Haseman & Gaylor	1002.47, 1063.63, 905.54	3.40, 0.75,43.13	4498.002	312.81
	Rubin	602.24, 1206.7, 447.16	403.63, 42.32, 501.51	12081.82	348.38
Y_{111}, $Y_{213}Y_{312}$	Coons	1016.72, 1295.38, 1207.296	10.84, 128.41, 44.81	3936.98	308.01
	Haseman & Gaylor	1016.72, 1295.38, 1207.296	10.84, 128.41, 44.81	3936.98	308.01
	Rubin	692.15, 1026.55, 920.85	313.72, 140.43, 241.64	12592.65	349.87
Y_{111}, $Y_{233}Y_{342}$	Coons	1016.72, 899.66, 1112.89	10.84, 89.04, 31.27	4256.628	310.82
	Haseman & Gaylor	1016.72, 899.66, 1112.89	10.84, 89.04, 31.27	4256.628	310.82
	Rubin	765.67, 625.19, 881.08	240.19, 363.50, 200.50	10659.55	343.87

*Note: * Y_{ijk} indicates the yield of i^{th} main plot in j^{th} sub plot in k^{th} replication, i = 1, 2, 3; j = 1, 2, 3, 4 and k = 1, 2, 3*

equivalently better than the method described by Rubin for all three criteria under study.

The above study firstly leads to the conclusion that the estimation and analysis of missing observations in experiments under split plot set up can be done successfully through ANCOVA model and the method can be used as an efficient alternative tool to estimate the missing observations in multi- factor experiments specially in case of experiments in split plot set up.

We also conclude that the three methods to estimate missing observations under consideration are equally efficient for estimation of single missing observation. But for multiple missing observations, the ANCOVA methods given by Coons (1957) and Haseman and Gaylor's method (1973) have equal precision and both are better than the method given by Rubin (1972).

REFERENCES

Ahmed, L. A. 2016. Missing Value Estimation Comparison in Split-Plot Design. *Int. J. Comp. Info. Tech.*, **5:** 337- 44.

Allan, F. E. and Wishart, J. 1930. A Method of Estimating the Yield of a Missing Plot in Field Experimental Work. *The J. Agril. Sci.*, **20:** 390- 406.

Anderson, R. L. 1946. Missing-Plot Techniques. *Biom. Bull.*, **2:** 41- 47.

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

Bartlett, M. S. 1937. Some Examples of Statistical Methods of Research in Agriculture and Applied Biology. *J. Royal Stat. Soc.*, **4:** 137- 83

Haseman, J. K. and Gaylor, D. W. 1973. An Algorithm for Non-iterative Estimation of Multiple Missing Values for Crossed Classifications. *J. Technomet.*, **15:** 631- 36.

Rubin, D. B. 1972. A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Appl. Stat.*, **21:** 136- 41.

Yates, F. 1933. The analysis of replicated experiments when the field results are incomplete. *Empire J. Exp. Agric.*, **1:** 129- 42.