International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 6, Issue 4, Jun – Jul 2017; 55-66

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

MISSING PLOT TECHNIQUES USING REGRESSION ANALYSIS AND ITS COMPARISON WITH RANDOMIZED BLOCK DESIGN MISSING PLOT TECHNIQUES

RICHA SETH[1], D. K. GHOSH[2] & N. D. SHAH[3]

[1]K. S. School of Business Management, Gujarat University, Ahmedabad, Gujarat, India

[2]Department of Statistics, Saurashtra University, Rajkot, Gujarat, India

[3]Principal, M. C. Shah Commerce College, Ahmedabad, Gujarat, India

ABSTRACT

In this paper, we have estimated the values of one and two missing observations using the regression method of analysis. Again, we estimated the value of same missing observations (one and two) by expressing the regression data in the format of randomized block design data. In this investigation, we observed that the estimated value of the missing observation(s) come out to be more or less same. This procedure has been shown by taking two suitable examples. Further, we analyzed the data to obtain ANOVA table using both the methods, where we found the same significant result.

KEYWORDS: Analysis of Variance, Blocking, Missing Observation, Regression Analysis, Response Variable

1. INTRODUCTION

For the statistical procedure in Design of Experiments, the matrix presentation of the data of a randomized complete block design is similar to that of the exhibit of a factorial experiment with two factors being studied at different levels. When there is a missing value in the dataset of a randomized block design, we obtain the estimate of missing value using least square method of estimation and then carry out the analysis of variance using the estimated value of the missing observation. Regression analysis provides a mathematical relationship between the response variable and the factors affecting it. The expression of relationship obtained by regression can also give an estimate of the missing value, so that the further analysis can be performed. A study of the relation between the Analysis of Variance and Regression Analysis techniques has been carried out by Arner [1] and Karen [5] in the context of data analysis. For the statistical analysis procedure, regression analysis, requirement of the data is in the form of response variable values and the values of the factors affecting it, without considering the levels of the factors involved. In a variation of this, the statistical procedure design of experiments requires the data in the form of response variable values being affected by different levels of the factors involved.

1.1. Analysis of Variance

The statistical tool Analysis of Variance is applied to analyze the data of a Design of experiment. Fisher [4] introduced the term ‘Analysis of Variance’ and defined it as the separation of variance ascribable to one group of causes from the variance ascribable to the other group. Under this technique, the total variation in the sampled data is divided into components of variation due to different independent factors. Each of these estimates of the variations due to assignable factors is compared with the estimate of the variation due to chance factor and identified whether the variation due to the assignable cause is significant or not.

1.2. Randomized Complete Block Design

If an experiment requires a large number of experimental units and all the experimental units are not homogeneous, then some measure of error control is needed. In a randomized block design, the error control measure is applied by dividing the experimental units into homogeneous groups and considering all the treatments to be studied in each group [2]. The groups are known as blocks, and the experimental units in the blocks are known as plots. Randomized block design is a complete block design where v treatments are arranged in b blocks, such that each block contains all the treatments once and each treatment is replicated r times.

2.1. Estimate of a Missing Observation in a Randomized Block Design

Yates [9] studied that the selection of the value as an estimate for a missing observation depends on the criterion of minimizing the error sum of squares [2]. ‘Suppose two observations are missing in a randomized block design, with k treatments and r replications. Let these missing observations belong to different blocks and affect different treatments and be substituted by the unknowns y_1 and y_2. Suppose y_1 belongs to the j^{th} block of i^{th} treatment and y_2 belongs to j' th block of m^{th} treatment. Let B_j denote the observation total of j^{th} block taking zero for the missing observations. Similarly, $B_{j'}$ denotes the total of the j' th block. T_i and T_m are similar totals for i^{th} and m^{th} treatments respectively. Further, let G denote the grand total of the observations obtained by taking zero for the missing observations.’

Incidentally, if only one observation is missing, its estimate can be obtained from the following equation.

$$y_1 = \frac{kT_i + rB_j - G}{(r-1)(k-1)} \quad \textbf{(1.1)}$$

If two observations are missing, then their estimates can be obtained from equations given as follows.

$$y_1 = \frac{(r-1)(k-1)(kT_i + rB_j - G) - (kT_m + rB_{j'} - G)}{(r-1)^2(k-1)^2 - 1} \quad \textbf{(1.2)}$$

$$y_2 = \frac{(r-1)(k-1)(kT_m + rB_{j'} - G) - (kT_i + rB_j - G)}{(r-1)^2(k-1)^2 - 1} \quad \textbf{(1.3)}$$

Once the missing observations are estimated using (1.2) and (1.3), we substitute these values at the missing plots and then analyse the data using ANOVA.

2.2. Regression Analysis

In Regression Analysis, a functional relationship between the response variable and the explanatory variables depicts how changes in the independent variables affect the values of the response variable [3]. The functional relationship is modelled on the assumption that some linear relationship between unknown parameters exists. Sometimes, linear relationship is not appropriate and non-linear model needs to be fitted. The relationship between the response variable and the explanatory variables also helps in estimating the unknown values and predicting the future values.

3. STUDY METHODOLOGY

In this paper, we have discussed and compared the estimates of the missing observations obtained by two procedures, namely, Design of Experiments and Regression Analysis. Further, we have also compared the analysis obtained by ANOVA and Regression Analysis after the missing observations are replaced by their respective estimates. The study is in the similar line of Arner [1] and Karen [5]. It is to be noted here that this study is done for the dataset whose format can be expressed interchangeably in both the forms: regression data as well as randomized block design. Moreover, it is to be noted that it would be cost effective if the regression data can be used for forming a layout of a designed experiment without actually conducting the experiment itself. This idea may also be applicable in using the regression data to estimate the observations which may be missing for certain combinations of the classified data in the context of Design of experiments.

Here to explain the concept, we took two examples which have been considered to justify the relationship between two statistical procedures, namely, Analysis of Variance and Regression Analysis. The first example, originally requires a regression model to be fitted to two variables A and B, but the data of the problem are such that it can be rewritten in the format of a designed experiment with three levels of factor A in the rows and two levels of factor B in the columns. Assuming one of the observations in the original data set to be missing and considering that the interaction among the factors is not present, the further analysis is conducted. The non-existence of interaction of the factors is identified by a regression model that is, fitted to the incomplete dataset using SPSS [8]. The fitted regression model is used to estimate the missing value. Next, the regression analysis is conducted on the dataset obtained after putting the estimate of the missing value.

Further, for the same dataset the randomized block design format of the data is considered. The missing value is again estimated using RBD and then ANOVA is performed on the dataset by replacing the missing observation with this estimate. The results of both the analyses, namely, ANOVA and Regression Analysis are compared. It is found that both the analyses of the dataset give similar conclusions.

The second example belongs to Design of experiment problem where ANOVA is to be performed. However, the data of the example are rewritten in the regression format also. Here, two of the observations are assumed to be missing, which are estimated by the RBD procedure of missing observation technique. The results of the ANOVA of the completed dataset are obtained after putting estimates of the missing values and are compared with the inference of the Regression Analysis. The Regression Analysis is conducted on the dataset, obtained after replacing the missing observations by the regression estimate of the observations. Here again, it is found that the ANOVA and the Regression Analysis of the completed dataset give similar conclusions, respectively.

Next we have shown a relationship between Analysis of Variance and Regression Analysis techniques.

4. EXAMPLE 1: [6]

The following data were collected to determine the relationship between two processing variables and hardness of a certain kind of steel. Let Y denotes Hardness (Rockwell 30-T), A denotes Copper Content (percent) and B denotes Annealing Temperature (degrees F).

Table 4.1: Data for Hardness of Steel Affected by two Processing Variables

Sr. No	y	A	B
1	78.9	0.02	1000
2	55.2	0.02	1200
3	80.9	0.10	1000
4	57.4	0.10	1200
5	85.3	0.18	1000
6	60.7	0.18	1200

Here, we assumed that the fourth value of *y* is missing, that is, the value of the observation corresponding to the level combination, copper content of 0.10 percent and annealing temperature 1200 degrees F in the original data is missing. This is shown in Table 4.2.

Table 4.2 Data (With Missing Observation) for Hardness of Steel Affected by Two Processing Variables

Sr. No	y	A	B
1	78.9	0.02	1000
2	55.2	0.02	1200
3	80.9	0.10	1000
4	*	0.10	1200
5	85.3	0.18	1000
6	60.7	0.18	1200

The following regression model is fitted to the data set of Table 4.2.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad \textbf{(4.1)}$$

It is to be noted here that we represent the factors A and B by the variables x_1 and x_2, respectively, in the regression analysis.

Equivalently, in the matrix form we have

$$y = X\beta + \varepsilon \quad \textbf{(4.2)}$$

where, $y = \begin{pmatrix} 78.9 \\ 55.2 \\ 80.9 \\ * \\ 85.3 \\ 60.7 \end{pmatrix}$ $X = \begin{pmatrix} 1\ 0.02\ 1000 \\ 1\ 0.02\ 1200 \\ 1\ 0.10\ 1000 \\ 1\ 0.10\ 1200 \\ 1\ 0.08\ 1000 \\ 1\ 0.08\ 1200 \end{pmatrix}$ $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$

ε is a random error such that $E(\varepsilon) = 0$, $V(\varepsilon) = \sigma^2$ and the $\{\varepsilon\}$ are uncorrelated.

We have used SPSS [8] to fit a regression model to the data given in Table 4.2. While running the regression analysis, interaction AB is also considered. Correlation between Y and A, Y and B, and Y and AB etc. are obtained and shown in Table 4.3. We have shown an ANOVA table in Table 4.4.

Table 4.3: Correlations

		Y	A	B	AB
Pearson Correlation	Y	1.000	.223	-.974	.099
	A	.223	1.000	.000	.988
	B	-.974	.000	1.000	.123
	AB	.099	.988	.123	1.000
Sig. (1-tailed)	Y	.	.359	.003	.437
	A	.359	.	.500	.001
	B	.003	.500	.	.422
	AB	.437	.001	.422	.

Table 4.4: ANOVA Table in Context of Regression Analysis with Missing Observation

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	676.875	1	676.875	55.535	.005
	Residual	36.565	3	12.188		
	Total	713.440	4			
2	Regression	712.277	2	356.139	612.712	.002
	Residual	1.162	2	.581		
	Total	713.440	4			

Coefficients of β and their standard errors for both the models are shown in Table 4.5. On using a stepwise regression procedure to fit the regression model, variable B is entered under model 1 while variables B and A, both are entered under model 2. However, no variables are removed.

Table 4.5: Coefficients

Model		Unstandardized Coefficients		t	Sig.
		β	Std. Error		
1	(Constant)	200.450	17.280	11.600	.001
	B	-.119	.016	-7.452	.005
2	(Constant)	196.731	3.804	51.722	.000
	B	-.119	.003	-34.125	.001
	A	37.188	4.765	7.804	.016

Here, we obtained the R^2 value for model 1 as 0.949 and the R^2 value for model 2 as 0.998. The fitted regression model for the dataset with a missing observation is approximated as follows:

$$\hat{y} = 196.731 + 37.188x_1 - 0.119x_2 \qquad (4.3)$$

Thus the estimated value of y for the missing combination A_1B_1 is

$$\hat{y}_{Regression} = 196.731 + 37.188 \times 0.10 - 0.119 \times 1200 = 57.65 \qquad (4.4)$$

This estimated value of the fourth observation is more or less same as shown in Table 4.1.

Next, we obtain the estimate of the missing observation, using the expression in (1.1) for a randomized block design. The same dataset can be rewritten as a two-way classified design as follows. Factor A is on three levels, namely, 0.02, 0.10 and 0.18. These levels are coded as 0, 1 and 2, respectively. Similarly, factor B is on two levels, namely, 1000 and 1200, which are coded as 0 and 1, respectively. Since in regression analysis of the data we found factor B to be highly significant, hence we consider factor B as a treatment and factor A as a block in a randomized block design. The RBD format of the data is shown in Table 4.6.

Table 4.6: Randomized Block Design Format of the Data with Missing Observation

Block (Factor A)

		0	1	2	$y_{i.}$
Treatment (Factor B)	0	78.9	80.9	85.3	245.1
	1	55.2	*	60.7	115.9
	$y_{.j}$	134.1	80.9	146	361

The estimate of the value of the missing observation can be obtained from

$$\hat{y}_{RBD} = \frac{kT_i + rB_j - G}{(r-1)(k-1)}$$

Here T_i, B_j and G are obtained by taking zero for the missing observation.

Considering Table 4.6, we have k = 2, r = 3, $T_2 = 115.9$, $B_2 = 80.9$ and G = 361. Thus,

$$\hat{y}_{RBD} = \frac{2\times115.9+3\times80.9-361}{2\times1} = 56.75 \quad \textbf{(4.5)}$$

From expressions (4.4) and (4.5) we found that the estimates of the missing observation obtained by regression analysis and RBD method are more or less similar. Now, we would analyze the completed dataset, first by putting the estimate of missing value given in (4.4) and shown in Table 4.7 by performing regression analysis and secondly, by putting the estimate of the same missing value given in (4.5) and shown in Table 4.10 by conducting ANOVA for RBD. Consequently, we have compared the Regression Analysis with ANOVA table.

Table 4.7 represents the data as regression format data after putting the missing value.

Table 4.7: Data for Hardness of Steel Affected by Two Processing Variables When Missing Observation is Replaced with Regression Estimate

Sr. No.	y	A	B
1	78.9	0.02	1000
2	55.2	0.02	1200
3	80.9	0.10	1000
4	57.65*	0.10	1200
5	85.3	0.18	1000
6	60.7	0.18	1200

The SPSS output of the regression analysis of the above dataset is as follows. The ANOVA; and the β coefficients and their standard errors are shown in the Tables 4.8 and 4.9 respectively.

Table 4.8: ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	853.234	1	853.234	93.186	.001
	Residual	36.625	4	9.156		
	Total	889.859	5			
2	Regression	888.636	2	444.318	1.090E3	.000
	Residual	1.223	3	.408		
	Total	889.859	5			

Table 4.9: Coefficients

Model		Unstandardized Coefficients		t	Sig.
		β	Std. Error		
1	(Constant)	200.950	13.645	14.727	.000
	B	-.119	.012	-9.653	.001
2	(Constant)	197.231	2.906	67.870	.000
	B	-.119	.003	-45.758	.000
	A	37.187	3.990	9.321	.003

From Table 4.9, we can say that both the factors A and B are significant for both the models 1and 2. It is to be noted here that the interaction AB is not even included in the model. When only variable B is entered in model 1, R^2 is 0.959, however, when variable A is also entered along with variable B in model 2, R^2 comes out as 0.999. The difference in R^2 value in both the models is 0.040, which is very small. So we can say that, in model 1 when factor B is present R^2 is 0.959, however, in model 2 when factor A entered with factor B, the R^2 increased by merely 0.040. Thus we can say that effect of factor A is much less as compared to factor B.

Next, we present the completed dataset in the RBD format so as to perform ANOVA.

Table 4.10: Randomized Block Design Data When Missing Observation is Replaced with RBD Estimate

		Block (Factor A)			
		0	1	2	$y_{i.}$
Treatment (Factor B)	0	78.9	80.9	85.3	245.1
	1	55.2	56.75*	60.7	172.65
	$y_{.j}$	134.1	137.65	146	417.75

Here $y_{..} = 417.75$ as the missing value is filled up by the estimated value 56.75 given in expression (4.5).

The model for the randomized block design is given by

$$y_{ij} = \mu + t_i + b_j + \varepsilon_{ij} \quad \textbf{(4.6)}$$

where, μ is general mean effect, t_i is the effect of the i^{th} treatment, b_j is the effect of the j^{th} block, ε_{ij}, the error component, are random variables assumed to be normally and independently distributed with mean zero and variance σ^2, i = 1, 2, …, k, j = 1, 2, …, r, where k is the number of treatments and r is the number of blocks. For the given example, we have k = 2 and r = 3.

The analysis of the RBD in Table 4.10 is given in the ANOVA Table 4.11.

Table 4.11: ANOVA Table for Hardness of Steel Affected by Two Processing Variables When Missing Observation is replaced with RBD Estimate

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	Variance Ratio (F_0)
Treatments (Factor B)	874.8337	1	874.8337	4320.17
Blocks (Factor A)	37.3225	2	18.6613	92.15
Error	0.2025	1	0.2025	
Total	**912.3587**	**4**		

Here, from Table 4.11, it is found that both the factors A (Copper Content) and B (Annealing Temperature) are significant in affecting the hardness of Rockwell 30-T steel. The factor A is significant while factor B is highly significant. The result is similar to what has been observed in regression analysis in Table 4.9.

Thus, we conclude that the result of Regression Analysis is the reflection of the inference drawn from the Analysis of Variance in context of the design of the experiment, which is completed by the estimated value. Here, it is being observed that Regression Analysis and Analysis of Variance are complementary to each other. As such regression model can be used to estimate the missing observations from a missing plot experiment and then usual analysis can be performed. To authenticate this observation, another example has been considered, where number of missing observations is two.

5. EXAMPLE 2: [7]

An industrial engineer is conducting an experiment on eye focus time. He is interested in the effect of the distance of the object from the eye on the focus time. Four different distances are of interest. He has five subjects available for the experiment. Because there may be differences between individuals, he decides to conduct the experiment in a randomized block design. The data obtained are shown below in Table 5.1.

Table 5.1: Randomized Complete Block Design for the Eye Focus Time Experiment

Distance (ft)	Subject				
	1	2	3	4	5
4	10	6	6	6	6
6	7	6	6	1	6
8	5	3	3	2	5
10	6	4	4	2	3

Here, it is assumed that the values of two observations corresponding to pairs (6, 1) and (10, 4) are missing, where in the bracket, the first digit shows the distance while the second digit shows the subject. We call them missing plot experiments. The data are given in the Table 5.2.

Table 5.2: Randomized Block Design (With Missing Observations) for the Eye Focus Time Experiment

Distance (ft)	Subject				
	1	2	3	4	5
4	10	6	6	6	6
6	y_1	6	6	1	6
8	5	3	3	2	5
10	6	4	4	y_2	3

Using equations (1.2) and (1.3), we estimate the missing values of the corresponding observations for the given RBD. The estimated values of the missing observations are $(\hat{y}_1)_{RBD} = 7.63\ and\ (\hat{y}_2)_{RBD} = 1.45$ **(5.1)**

Substituting these estimates for the missing values and performing ANOVA on the completed dataset, the following result is obtained and is shown in Table 5.3.

Table 5.3: Analysis of Variance of Missing Observations Experiment in Context of Design of Experiment

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Sum of Squares	Variance Ratio (F_0)
Distance	34.6856	3	11.5619	7.75
Subject	41.4936	4	10.3734	6.96
Error	14.9139	10	1.4914	
Total	**91.0931**	**17**		

Viewing Table 5.3 it is observed that both treatment effects and block effects are significant. It is found that the four different distances of the object from the eye lead to different focus times for the eye.

The above analysis can be compared with the results of the regression analysis conducted on the same data rewritten in the following regression format, shown in Table 5.4.

Table 5.4: Data (With Missing Observations) for the Eye Focus Time Experiment

Sr. No	y	Treatments	Blocks
1	10	4	1
2	6	4	2
3	6	4	3
4	6	4	4
5	6	4	5
6	y_1	6	1
7	6	6	2
8	6	6	3
9	1	6	4

Table 5.4: Contd.,

Sr. No	y	Treatments	Blocks
10	6	6	5
11	5	8	1
12	3	8	2
13	3	8	3
14	2	8	4
15	5	8	5
16	6	10	1
17	4	10	2
18	4	10	3
19	y_2	10	4
20	3	10	5

The regression analysis of the data is performed with SPSS. We found that the Stepwise regression procedure does not include the subject variable in the model. In Stepwise method inclusion or removal of one independent variable is done at each step, based (by default) on the probability of F (p-value). The default probability of F for entering the variable is 0.05 and that for removing the variable is 0.10. So, when we tried to fit the regression model by Stepwise procedure subject variable is not entered in the model. Also, the R^2 value of this model is very low i.e. 0.277. So we tried to fit the regression model with the Enter regression procedure. In Enter regression method all independent variables are entered into the equation in one step. This led to a better regression model with an R^2 value as 0.425. The fitted regression model using the Enter method is shown in Table 5.5 and Table 5.6.

Table 5.5: ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30.474	2	15.237	5.534	.016
	Residual	41.304	15	2.754		
	Total	**71.778**	**17**			

Table 5.6: Coefficients

Model		Unstandardized Coefficients		t	Sig.
		β	Std. Error		
1	(Constant)	10.027	1.595	6.288	.000
	Distance	-.500	.176	-2.847	.012
	Subject	-.553	.282	-1.964	.068

It is to be noted here that the regression analysis is performed on the missing data and it is observed that subject effects seem to be insignificant at 5% level of significance. It would be interesting to note the results when the regression is performed on the completed dataset by putting the regression estimates of the missing values. The expression of the regression model for the missing observation dataset is as follows:

$$\hat{y} = 10.027 - 0.500x_1 - 0.553x_2 \quad \textbf{(5.2)}$$

Here, in the regression analysis, the representation of the distance variable is done by the variable x_1 and that of the subject variable is done by the variable x_2.

The estimated values of the missing observations using the regression model are:

$$(\hat{y}_1)_{Reg} = 6.474 \ and \ (\hat{y}_2)_{Reg} = 2.815 \quad \textbf{(5.3)}$$

From expressions (5.1) and (5.3) it is seen that the estimates of the missing observations obtained from randomized block design method and regression analysis techniques, respectively, are nearby the actual values of the observations, i.e. $y_1 = 7\ and\ y_2 = 2$.

The estimated values obtained from regression analysis are now substituted in place of missing observations and regression analysis is performed on the complete dataset. The analysis is given as follows in the Tables 5.7 and 5.8.

Table 5.7: ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.029	1	25.029	8.413	.010
	Residual	53.550	18	2.975		
	Total	78.579	19			
2	Regression	37.276	2	18.638	7.671	.004
	Residual	41.304	17	2.430		
	Total	**78.579**	**19**			

Table 5.8: Coefficients

Model		Unstandardized Coefficients		t	Sig.
		β	Std. Error		
1	(Constant)	8.366	1.267	6.601	.000
	Distance	-.500	.172	-2.901	.010
2	(Constant)	10.026	1.363	7.354	.000
	Distance	-.500	.156	-3.210	.005
	Subject	-.553	.246	-2.245	.038

From Table 5.8, it can be viewed that using the Stepwise regression procedure only distance variable is entered under model 1 while under model 2 subject variable is also entered along with distance variable. For model 1, the R^2 value is 0.319 and for model 2, the R^2 value is 0.474. Moreover, the distance effects and subject effects both are significant. Thus the four different distances of the object from the eye lead to different focus timings of the eye. This inference is similar to what we obtained in the Analysis of Variance of the RBD with missing observations filled by the RBD estimates.

Remarks

We have noticed here that the analysis of variance of the dataset obtained after replacing the missing values with the RBD estimates, given in the ANOVA Table 5.3 and the above regression analysis of the dataset obtained after replacing the missing values with the regression estimates, given in Table 5.8, both converge to the same results. Distance and subject effects are found to be significant. Thus, it creates confidence in using a regression model to estimate the observations which may be missing for certain plots in a randomized block design experiment. This idea may be applied for using the regression data for forming a layout of a designed experiment (if the format of the data permits), without actually conducting the experiment itself.

6. CONCLUSIONS

Thus, regression analysis is shown to be helpful in filling up the missing data in a randomized block design experiment and carrying out the analysis. Moreover, the datasets of randomized block design and regression format can be interchangeably used for analysis purposes which would be cost effective procedure for quality improvement.

REFERENCES

1. Arner M. (2014), Statistical Robust Design: An Industrial Perspective, John Wiley and Sons.
2. Das M. N. and Giri N. C. (1979), Design and Analysis of Experiments, Wiley Eastern Limited, New Delhi.
3. Draper N. R. and Smith H. (1966), Applied Regression Analysis, 2nd ed., John Wiley and Sons, USA.
4. Fisher R. A. (1925), Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh, London.
5. Karen G. M. (2016), The Analysis Factor - Making Statistics Make Sense, available at http://www.theanalysisfactor.com/why-anova-and-linear-regression-are-the-same-analysis/.
6. Millier I. and Millier M. (2004), John E. Freund's Mathematical Statistics with Applications, 7th ed., Pearson Education, Delhi, 480.
7. Montgomery, D. C. (1976), Design and Analysis of Experiments, 1st ed., John Wiley and Sons, New York, 96.
8. SPSS 21, IBM SPSS Statistics, Version 21.
9. Yates, F. (1937), The design and analysis of factorial experiments, Technical Communication no. 35 of the Commonwealth Bureau of Soils (alternatively attributed to the Imperial Bureau of Social Science).