African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 54-67)
www.abjournals.org

COMPARISON OF METHODS OF ESTIMATING MISSING VALUES IN RANDOMIZED COMPLETE BLOCK DESIGNS (RCBD) FOR VARIOUS NUMBER OF MISSING VALUES

Nwankwo Chike H and Onah Jude Chinedu

Department of Statistics, Nnamdi Azikiwe University, Awka. Nigeria

ABSTRACT: *In virtually all types of research, missing value(s) are common occurrences. Several methods and techniques are available for handling this issue. This study aims at comparing the result of using some of the techniques (Do nothing, List-wise Deletion, Overall Mean Imputation, Group Mean Imputation and Least squares Imputation techniques) in handling missing values in Randomized Complete Block Designs (RCBDs). Research data on outcome of an experiment conducted in July 2010 in the department of Animal Science, University of Nigeria Nsukka was employed. The data is made up of a random sample of size 36, on the effect of stocking Densities on the weight of birds at varying ages. Weight gain was used as parameter for measurement. Missing value(s) were introduced into the original complete data set randomly in three different levels of n = 1, 2 and 3. A two-way analysis was carried out, when the data is complete and when there are missing value(s) using the different methods of handling missing values considered. Results showed that the model assumptions of the RCBD was the same, both when the data is complete and when using the different methods of handling missing value(s) employed at different levels of n = 1,2 and 3 missing value(s). There is significant effect of the stocking densities when the data is complete and when using the different method of handling missing value(s) at different levels of n = 1, 2 and 3 missing values. On the contrary, there are significant differences between the MSE of the analysis with the complete dataset and when the different methods of handling missing value(s) at different levels (n=1, 2 and 3) of missingness are created. The complete data showed an MSE of 0.538. On the other hand, for n=1, 2 and 3 missing values respectively, the "Do Nothing" technique generated MSEs of 0.559, 0.573 and 0.577 (average 0.570). Listwise Deletion showed MSEs of 0.564, 0.534 and 0.714, (average 0.604). Random Mean Imputation showed MSEs of 0.537, 0.629 and 0.776, (average 0.657). Group Mean Imputation generated MSEs of 0.553, 0.704 and 0.785, (average 0.681). Least Squares Imputation produced MSEs of 0.553, 0.527 and 0.527, (average 0.530). Hence, the Least Squares Imputation, with consistently small MSEs and the closest average MSE to the true MSE, is recommended among the methods studied.*

KEYWORDS: Design, Orthogonal, Randomize, Missingness, Complete, Block

INTRODUCTION

Missing observations is a regular occurrence in many researches works, even in well-planned experiment. In analyzing data generated from the Randomized Complete Block Designs (RCBDs) , missing vlues abound also. As a result of the orthogonal nature of the original structure of the RCBD, RCBD has various applications both in Agriculture, Engineering, Medicine, Pharmaceutical Sciences, etc. Klaus and Oscar (2007) stated that the simplest and perhaps the most widely used block design is the Randomized Complete Block Design

(RCBD). In any research work that involves data generation, there is always a possibility of experiencing missing values. Causes of missing values are numerous, some are; high cost of research, limited time available for the research, faulty apparatuses, death or incidental loss of experimental units, error due to parallax, when experiment is being carried out by quacks, geographical locations, carelessness etc. Ella, (2006) stated that even in a well-planned experiment, using RCBD, it may happen that for reasons that may not be ascribed to the effect of the treatments, and beyond the control of the experimenter, one or several observations may not be available. When this occurs in an experiment that has all the features of RCBD, it is said that there exists missing observation(s). According to Douglas (2000), when missing values occur in RCBD, the design becomes an incomplete block design which leads to deviation from the standard statistical analyses of the RCBD, having lost its orthogonality nature and becomes Incomplete Block Design. Jarett, (1978) suggested that a reasonable solution to this problem would be to insert values for the missing observations and to perform the analysis using the structure of the original experiment called the imputation technique.

There are several imputation techniques and other model-based approaches for handling missing value(s) in RCBD as well as other types of designs of experiments. Some of these techniques includes Blockwise Mean Imputation (Group Mean Imputation), Random Mean Imputation, Listwise Deletion, Do-nothing Technique or Complete-Case Analysis, Least Squares Imputation, Hot-Deck Imputation, Expectation Maximization and Nearest-Neighbor Imputation.

The aim of this paper is to compare and recommend method or methods of imputation of missing observation(s) in Randomized Complete Block Designs (RCBD) as the number of missing values n are 1, 2 or 3.

In RCBDs, missing values cause non orthogonality of the design, hence loss of precision of the inferences involved, Chakrabarti (1962).

Because there are many types of data realizations, different levels of missingness, different types of missing values and hence different types of imputation techniques, a "one measure fits all" decision does not appear to suffice, that is, one cannot say for sure, the best imputation technique for handling all cases of missing values in all sets of data. The effort here is to determine the most approximate imputation technique at various levels of missingness for $n \leq 3$ missing values in RCBD.

In this paper, Least Squares Imputation, Random Mean Imputation, Blockwise Mean Imputation, Listwise Deletion and Do-Nothing Techniques for the imputation of missing observations in the analysis of RCBD will be compared.

METHODOLOGY

Data for this research was collected from the Department of Animal Science, University of Nigeria, Nsukka. The data is the outcome of the experiment conducted by a group of Animal Scientists, in that Department, on the Effect of Stocking Densities on the Weight of Birds at Varying Ages. The experiment was conducted around July 2010. Table 1 below shows the results from the experiment.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 54-67)

www.abjournals.org

Table 1: Average Weight Gains by Birds for Varying Stocking Densities at Different Ages

	Treatment (Stocking Density)					
Block (Age in Weeks)	**T_1 0.9m²/ Bwd**	**T_2 0.41m²/ Bwd**	**T_3 0.27m²/ Bwd**	**T_4 0.25m²/ Bwd**	**T_5 0.21m²/ Bwd**	**T_6 0.2m²/ Bwd**
b_1(16)	9.07	9.07	8.54	8.32	7.99	7.92
b_2(17)	9.52	9.16	9.37	8.26	9.03	10.02
b_3(18)	10.54	10.01	9.93	9.04	9.41	11.24
b_4(19)	12.18	10.96	11.03	12.41	10.52	10.19
b_5(20)	12.71	11.56	11.45	12.51	11.97	13.92
b_6(21)	13.53	13.12	11.84	12.09	13.01	11.67

For the purpose of this paper, the different stocking densities represent the treatment factors while the different ages of the birds, in weeks, form the blocking factors. The experiment was carried out for six weeks.

At the end of the experiment, the average weight gain of the birds was observed and recorded. The data are made up of samples of size 36 (i.e. 6 birds per cell).

Model Adequacy Check/Verification of Model Assumptions

Model assumptions of the RCBD include Independence among the data values, Normality assumption of the Error Terms, Constant Variance of the Error Terms, and Additivity Criterion, Cochran and Cox, (1957). Relevant computations were carried out using the R-Software.

When testing for normality the Normal probability plot and the Shapiro-wilks test for normality of residuals were computed.

Analysis on the Original Complete Research Data

To tests for conformity of the original data to the assumptions of the RCBD, the following results would be obtained (Klaus *et al,* 2018)

$$\hat{y}_{ij} = \bar{y}_{i} + \bar{y}_{.i} - \bar{y}_{..} \qquad [1]$$

$\hat{y}_{ij}$ = the predicted value and

Also

$$\bar{y}_{..} = \frac{\sum_{i=1}^{u} y_{i.}}{N} = \frac{\sum_{j=1}^{v} y_{.j}}{N}$$

$$\bar{y}_{i.} = \frac{\sum_{j=1}^{v} y_{ij}}{v}$$

$$\bar{y}_{.j} = \frac{\sum_{i=1}^{u} y_{ij}}{u}$$

i = 1… u

j = 1… v

Residuals would be generated using equation 2 thus

$\epsilon_{ij} = y_{ij} - \hat{y}_{ij}$ [2]

y_{ij} = the observed value

The results obtained are presented in table 2 below.

Table 2: Observations, Predictions, Residuals and Standardized Residuals from the Original Complete Data

S/N	Observations	Prediction	Residual	Standardized Residual
1	9.07	9.1014	-0.0314	-0.0514
2	9.52	9.8431	-0.3231	-0.5287
3	10.54	10.6447	-0.1047	-0.1714
4	12.18	11.8314	0.3486	0.5705
5	12.71	12.9697	-0.2597	-0.4250
6	13.53	13.1597	0.3703	0.6059
7	9.07	8.4897	0.5803	0.9496
8	9.16	9.2314	-0.0714	-0.1168
9	10.01	10.0331	-0.0231	-0.0377
10	10.96	11.2197	-0.2597	-0.4250
11	11.56	12.3581	-0.7981	-1.3060
12	13.12	12.5481	0.5719	0.9359
13	8.54	8.2031	0.3369	0.5514
14	9.37	8.9447	0.4253	0.6959
15	9.93	9.7464	0.1836	0.3005
16	11.03	10.9331	0.0969	0.1586
17	11.45	12.0714	-0.6214	-1.0169
18	11.84	12.2614	-0.4214	-0.6896
19	8.32	8.2814	0.0386	0.0632
20	8.26	9.0231	-0.7631	-1.2487
21	9.04	9.8247	-0.7847	-1.2841
22	12.41	11.0114	1.3986	2.2887
23	12.51	12.1497	0.3603	0.5896
24	12.09	12.3397	-0.2497	-0.4087
25	7.99	8.1647	-0.1747	-0.2859
26	9.03	8.9064	0.1236	0.2023
27	9.41	9.7081	-0.2981	-0.4877
28	10.52	10.8947	-0.3747	-0.6132
29	11.97	12.0331	-0.0631	-0.1032
30	13.01	12.2231	0.7869	1.2878
31	7.92	8.6697	-0.7497	-1.2269
32	10.02	9.4114	0.6086	0.9960
33	11.24	10.2131	1.0269	1.6805
34	10.19	11.3997	-1.2097	-1.9796
35	13.92	12.5381	1.3819	2.2615
36	11.67	12.7281	-1.0581	-1.7314

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 54-67)
www.abjournals.org

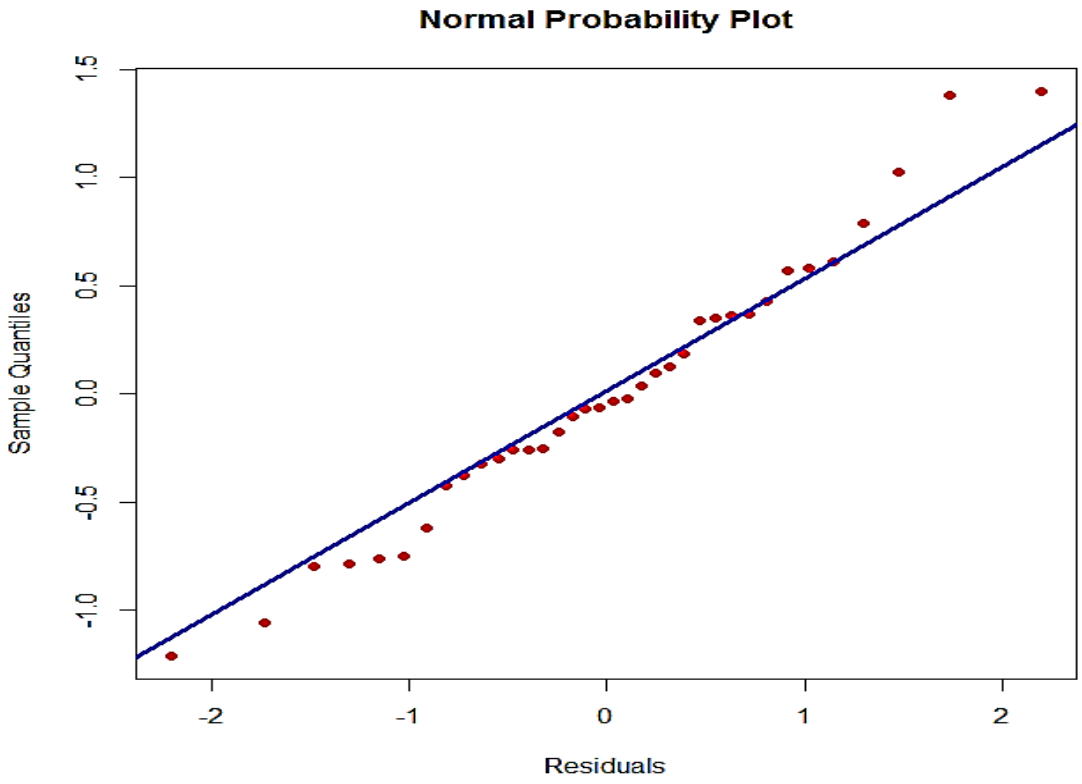

Figure 1: Normal Probability plot from the Original Complete Data

From the normal probability plot above, there is strong evidence of normality of the error terms. This is because the plotted points show little deviation from the straight line.

Shapiro-Wilk Test for Normality

To further confirm that the error terms are normally distributed, the Shapiro-Wilk test for normality is applied to the error terms obtained as shown in table 2, using he Shapiro-wilk test statistic according to Shapiro and Wilk, (1965),

$$W=\frac{(\sum_{i=1}^{n} a_i y_{(i)})^2}{(\sum_{i=1}^{n}(y_i-\bar{y})^2)} \quad [3]$$

Where

y_i = sample observations for i =1, 2, …, n

$y_{(i)}$ = the ordered sample observations

$\bar{y}$ = the mean of the observed sample

a_i, according to Sarhan and Greenberg (1956), are just the normalized best linear unbiased coefficients

Thus, we have

W = 0.97927, p-value = 0.7209

From the Shapiro-wilk table, the p-value corresponding to W = .97927 is 0.7209, hence the null hypothesis (H_0) that the errors are normally distributed is not rejected since the P-value = 0.7209 is greater than the level of significance, 0.05, thus it is concluded that the error terms are normally distributed.

Independence of the error terms is ensured by the proper randomization, at the design stage, being a randomized complete block design (RCBD).

Testing for homogeneity of variances, the Levenes test for homogeneity of variances is applied on the treatment factor. Given six (6) g groups of treatments, each normally distributed with different means and standard deviations σ_1, σ_2,…, σ_g. Also, given that n_1, n_2, …, n_g denote the number of six (6) subjects in each group of treatments, Y_{ki} denote response values and N denote the total sample size of thirty six (36) of all the six (6) groups, Levene (1960). The Levene's test statistic is given by

$$W=\frac{(N-g)\sum_{k=1}^{g} n_k (Z_k-\bar{Z})^2}{(g-1)\{\sum_{k=1}^{g}\sum_{i=1}^{n_k}(Z_{ki}-\bar{Z}_k)^2\}} \quad [4]$$

Where

$Z_{ki}=|Y_{ki}-\bar{Y}_k|$

$\bar{Z}_k=\frac{1}{n_k}\sum_{i=1}^{n_k} Z_{ki}$

$\bar{Z}=\frac{1}{N}\sum_{k=1}^{g}\sum_{i=1}^{n_k} Z_{ki}$

$\bar{Y}_k=\frac{1}{n_k}\sum_{i=1}^{n_k} Y_{ki}$

Here, the null and alternative hypotheses are

H_0: The treatment variances are all equal

H_1: The treatment variances are not all equal

The probability of F-value is the P-value from the output result.

If P(F-value) = P-value is less than alpha, the level of significance, we reject the null hypothesis H_0 and if the P(F-value) = P-value is greater than alpha, the level of significance, H_0 is accepted, hence conclude that variances are equal within treatments (Douglas, 2000).

Testing for Additivity of the factor variables in the RCBD model was carried out with the Turkey 1-df test for non-additivity. This assumption is barely necessary since the research data is made up of a single replication per block-treatment combination, that is, it is possible to directly measure the error or noise in this research. Turkey 1-df Test is to make sure the

block-treatment interaction is not significant and that the blocks are behaving as blocks so that the Mean Squared Error (MSE) can be used as a good estimator of the true experimental error Douglas, (2000).

The associated statistic for the Tukey's 1df test for non-additivity (John *et al*, 1984) is

$$F = \frac{\left[\sum_{i=1}^{u}\sum_{j=1}^{v} r_{ij}\epsilon_{ij}\right]^2 [(u-1)(v.1)-1]}{\sum_{i=1}^{u}\sum_{j=1}^{v} \epsilon_{ij}{}^2 - \left[\sum_{i=1}^{u}\sum_{j=1}^{v} r_{ij}\epsilon_{ij}\right]^2} \sim F^{1}_{(u-1)(v-1)-1}$$

Where

$$r_{ij} = \frac{\int(\bar{y}_{i.}-\bar{y}_{..},\bar{y}_{.j}-\bar{y}_{..})}{\sqrt{\sum_{i=1}^{u}\sum_{j=1}^{v}\iint(\bar{y}_{i.}-\bar{y}_{..},\bar{y}_{.j}-\bar{y}_{..})}}$$

$\epsilon_{ij} = y_{ij}\text{-}\bar{y}_{i.}\text{-}\bar{y}_{.j}+\bar{y}_{..}$

u = number of treatment

v = number of blocks

The hypotheses are,

H_O: The model effects are additive

Vs

H_1: The model effects are not additive

If the P-value from the output of the analysis is greater than the alpha value, H_0 is not rejected thus one would be justified in using the MSE which contains the block-treatment interaction as a reliable estimate of the true experimental error. A plot of the residuals versus the predicted values was also examined.

Generating ANOVA Table

Analysis of variance (ANOVA) table for the randomized complete block design for the existing data set will be generated for this research. The structure of the ANOVA table for the RCBD (Kutner *et al*, 2005), given the model;

$$Y_{ij} = \mu + t_i + b_j + Z_{ij} \qquad [5]$$

í = 1, 2, …, v

j = 1, 2, …, u

Subject to

$\sum_{i=1}^{v} t_i = \sum_{j=1}^{u} b_j = 0$

$Z_{ij} \sim N(0,\sigma_z{}^2)$

is shown in table 3 below

Table 3: Structure of ANOVA Table

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F-Ratio/p-value
Treatment	$SS_{Treatment}$= $\frac{1}{v}\sum_{i=1}^{u} y_{i.}{}^2 - \frac{y_{..}{}^2}{N}$	u-1	$\frac{SS_{Treatment}}{u-1}$	$\frac{MS_{Treatment}}{MS_{Error}}$
Blocks	SS_{Blocks} = $\frac{1}{u}\sum_{j=1}^{v} y_{.j}{}^2$ - $\frac{y_{..}{}^2}{N}$	v-1	$\frac{SS_{Block}}{v-1}$	
Error	SS_{Error} = SS_{Total}- $SS_{Treatment}$-SS_{Block}	(u-1)(v-1)		
Total	SS_{Total}= $\sum_{i=1}^{u}\sum_{j=1}^{v} y_{ij}{}^2$ - $\frac{y_{..}{}^2}{N}$	N-1		

Where $y_{i.} = \sum_{j=1}^{v} y_{ij}$, i = 1, 2, …, u

$y_{.j} = \sum_{i=1}^{u} y_{ij}$, j = 1, 2, …, v

$y_{..} = \sum_{i=1}^{u}\sum_{j=1}^{v} y_{ij} = \sum_{i=1}^{u} y_{i.} = \sum_{j=1}^{v} y_{.j}$

Testing for the Effect of Treatment Factor

This category involves building hypotheses to test for the significance of the effect of the treatment factor. The significance of this test is that it will form part of the basis for the comparison of the performances of the different methods of estimating missing values in Randomized Complete Block Designs.

For testing the effect of the treatment factor, the hypothesis is as follows;

H_0: $t_1 = t_2 = \ldots = t_3 = 0$

Vs

H_1: $t_i \neq 0$ for at least one i

The decision in support of any of the hypothesis above will be taken using the p-value from the ANOVA table (Table 1) at 0.05 level of significance. H_0 will be rejected if the P-value is less than alpha at 0.05, but will be accepted otherwise.

Creating/Introducing Missing Value(s)

The idea behind this research is on randomly deleting value(s) from the original complete data set to create missing value(s) scenarios. Through use of random Sampling without replacement, the process is repeated for n =1, 2, 3 and n missing value(s) and a discrete random variable. n=1,2,3 is simply a space chosen for the purpose of this research work

towards a robust comparison of methods of estimating missing values in RCBD. These actions generated the following tables:

Table 4: Research Data with One Missing Value

	Treatment (stocking density)					
Block (Age in Weeks)	**T_1 0.9m^2/Bwd**	**T_2 0.41m^2/Bwd**	**T_3 0.27m^2/Bwd**	**T_4 0.25m^2/Bwd**	**T_5 0.21m^2/Bwd**	**T_6 0.2m^2/Bwd**
b_1(16)	9.07	9.07	8.54	8.32	7.99	7.92
b_2(17)	9.52	9.16	9.37	8.26	9.03	10.02
b_3(18)	**NA**	10.01	9.93	9.04	9.41	11.24
b_4(19)	12.18	10.96	11.03	12.41	10.52	10.19
b_5(20)	12.71	11.56	11.45	12.51	11.97	13.92
b_6(21)	13.53	13.12	11.84	12.09	13.01	11.67

Table 4 shows the data table with one randomly chosen value omitted (represented as NA).

Table 5: Research Data with Two Missing Values

	Treatment (stocking density)					
Block (Age in Weeks)	**T_1 0.9m^2/Bwd**	**T_2 0.41m^2/Bwd**	**T_3 0.27m^2/Bwd**	**T_4 0.25m^2/Bwd**	**T_5 0.21m^2/Bwd**	**T_6 0.2m^2/Bwd**
b_1(16)	9.07	9.07	8.54	8.32	7.99	7.92
b_2(17)	9.52	9.16	9.37	8.26	9.03	10.02
b_3(18)	**NA**	10.01	9.93	9.04	9.41	11.24
b_4(19)	12.18	10.96	11.03	12.41	10.52	10.19
b_5(20)	12.71	11.56	11.45	12.51	11.97	13.92
b_6(21)	13.53	13.12	**NA**	12.09	13.01	11.67

Table 6: Research Data with Three Missing Values

	Treatment (stocking density)					
Block (Age in Weeks)	**T_1 0.9m^2/Bwd**	**T_2 0.41m^2/Bwd**	**T_3 0.27m^2/Bwd**	**T_4 0.25m^2/Bwd**	**T_5 0.21m^2/Bwd**	**T_6 0.2m^2/Bwd**
b_1(16)	9.07	9.07	8.54	8.32	**NA**	7.92
b_2(17)	9.52	9.16	9.37	8.26	9.03	10.02
b_3(18)	**NA**	10.01	9.93	9.04	9.41	11.24
b_4(19)	12.18	10.96	11.03	12.41	10.52	10.19
b_5(20)	12.71	11.56	11.45	12.51	11.97	13.92
b_6(21)	13.53	13.12	**NA**	12.09	13.01	11.67

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 54-67)
www.abjournals.org

Table 7: ANOVA Table of Levene's Test for Treatment Homogeneity for Complete Research Data

	Df	F value	Pr(>F)
Group	5	0.7295	0.6069
	30		

From the Levene's test table 7 above, we fail to reject the null hypothesis (H_0) that variances among treatments are homogeneous since P-value = 0.7273 is greater than the level of significance which is 0.05 and as a result, we conclude that variances among treatments are homogenous.

Table 8: ANOVA Table for Additivty Test for Complete Research Data

	df	**Sum of squares**	**Mean square**	**F-Value**	**Pr(F)**
Treatments	5	3.83	0.765	1.369	0.271
Blocks	5	83.43	16.685	29.857	1.47e-.09***
Square Prediction	1	0.03	0.031	0.056	0.816
Residuals	24	13.41	0.559		

*Significance Codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '*

From Table 8, we fail to reject the null hypothesis (H_0) that the model effects are additive, since the P-value = 0.816 is greater than the level of significance of 0.05, as a result, there is a justification in using the MSE which contains Block and treatment interaction (Block*Treatment) as a reliable estimate of the true experimental error.

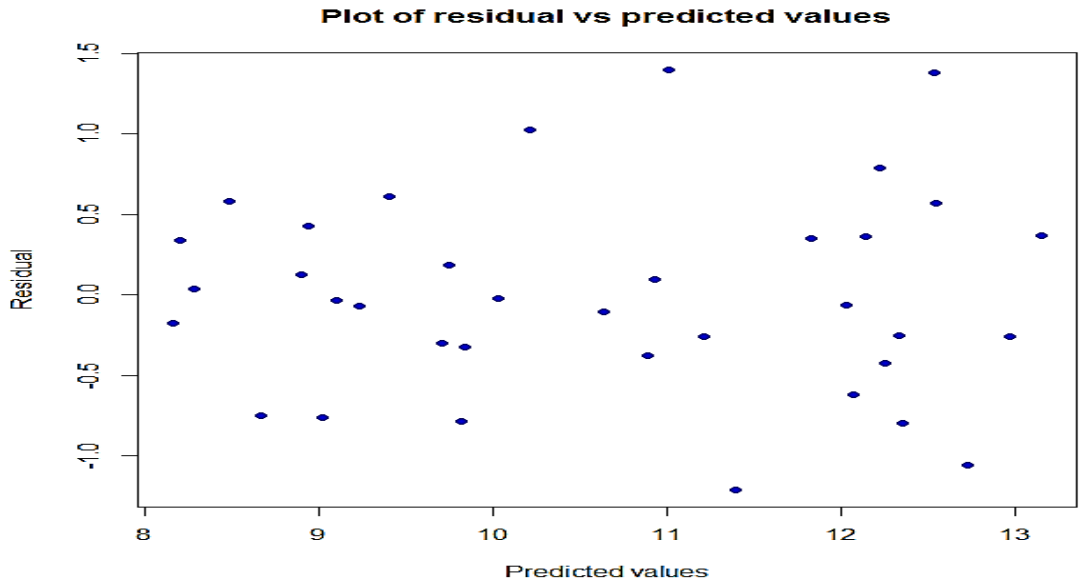

Figure 2: Plot of Residuals versus the Predicted Values of the Original Complete Research Data.

From figure 2, there is no evidence that there is any serious problem of nonadditivity or variance homogeneity among treatments. This was seen from the scattered nature of the plotted points.

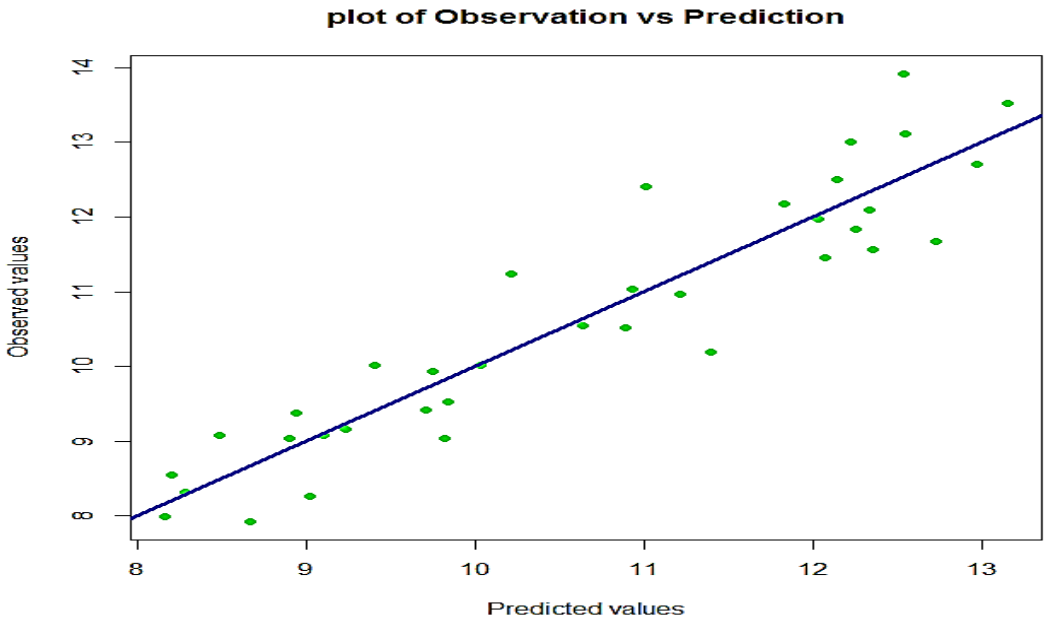

Figure 3: The Plot of Observed Values versus the Predicted Values of the Original Complete Research Data.

From figure 3, the plot of observed values versus the predicted values shows the relevance of this research data in future predictions due to the linear relationship that exist between the observed values and the predicted values.

Table 9: ANOVA Table of the Original Complete Research Data.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Treatment	5	3.83	0.765	1.423	0.25
Blocks	5	83.43	16.685	31.029	5.93e-10 ***
Residuals	25	13.44	**0.538**		

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

From table 9, since the P-value =0.25 is greater than the level of significance which is 0.05 we fail to reject the null hypothesis for the effectiveness of the treatments effect and accept that the treatment effect is effective.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 54-67)

www.abjournals.org

RESULTS

For each of the imputation techniques considered at three (3) different levels of missingness, same processes of analyses were carried out as for the Complete Research Data, making a total of sixteen different analyses. The results are which are summarized on the following tables.

Table 10: A Summary Table for the Analysis of the Research Data without any Missing Value

	Treatment Effects	Homogeneity of Variance of Error Terms	Normality of Error Terms	Additivity of Model Effects	MSE
Original Complete Data	Treatment effects are effective	Variance of the error terms are homogeneous	Error terms are normally distributed	Model effects are additive	0.538

Table 11: Summary Table for the Analyses of the Research Data with Missing Value(s) and when Applying Different Methods of Handling Missing Value(s) At Different Levels of Missingness

Methods of Handling Missing Values	**Treatment Effects**			**Homogeneity of Variance of Error Terms (Treatment wise)**			**Additivity of Model Effects**		
	n=1	n=2	n=3	n=1	n=2	n=3	n=1	n=2	n=3
Do Nothing Technique	Effective	Effective	Effective	Homogeneity	Homogeneity	Homogeneity	Additive	Additive	Additive
Listwise Deletion Technique	Effective	Effective	Effective	Homogeneity	Homogeneity	Homogeneity	Additive	Additive	Additive
Random Mean Imputation	Effective	Effective	Effective	Homogeneity	Homogeneity	Homogeneity	Additive	Additive	Additive
Group Mean Imputation	Effective	Effective	Effective	Homogeneity	Homogeneity	Homogeneity	Additive	Additive	Additive
Least Square Imputation	Effective	Effective	Effective	Homogeneity	Homogeneity	Homogeneity	Additive	Additive	Additive

Table 11A: Continuation of Table 11.

	Normality of error terms			MSE			Average
Methods of Handling Missing Values	**n=1**	**n=2**	**n=3**	**n=1**	**n=2**	**n=3**	
Do Nothing Technique	Normally Distributed	Normally Distributed	Normally Distributed	0.557	0.573	0.577	0.570
Listwise Deletion Technique	Normally Distributed	Normally Distributed	Normally Distributed	0.564	0.534	0.714	0.604
Random mean imputation	Normally Distributed	Normally Distributed	Normally Distributed	0.537	0.659	0.776	0.657
Group Mean Imputation	Normally Distributed	Normally Distributed	Normally Distributed	0.553	0.704	0.785	0. 681
Least Squares Imputation	Normally Distributed	Normally Distributed	Normally Distributed	0.537	0.527	0.527	0.530

CONCLUSION

From table 11 and 11A, the following conclusions are visible

- For n=1 missing value, the Group Mean imputation technique and the Least Squares imputation technique have the joint least MSE values of 0.537. Either of them is recommended for cases of 1 missing value in the RCBD.
- For n=2 missing values, the least Squares imputation technique has the least MSE of 0,527. The Least Squares imputation technique is therefore recommended for cases of 2 missing values in the RCBD.
- In the case of 3 missing values, the Least Squares imputation technique, again, showed superiority with the least MSE value of 0.527. It is thus recommended, for cases of 3 missing values in the RCDB,
- Finally, from the average MSEs, the Least Squares imputation technique showed the least MSE value of 0.530; this value is also closest to the true MSE for the complete dataset (0.538). It also showed smallest variation among all other techniques considered. With these attributes, for a researcher who experienced more than 3 missing values, the Least Squares imputation technique holds the potential of performing better than most other alternatives because of the consistent low and less variant MSEs, therefore the Least Squares is recommended for all cases of missing values imputation. The Group Mean imputation technique is recommended as a candidate only for cases of 1 missing value.

REFERENCES

Chakrabarti M.C. (1962). Mathematics of Design and Analysis of Experiments; London, Asia Publishing House.

Chigbu, P. E. (1998). Block Designs, Efficiency Factors and Optimality Criteria for Comparison; Enugu Nigeria, Linco Press Ltd.

Cochran, W. G. and Cox, G. M. (1957). Experimental Designs, 2nd edn; New York: Wiley.
Cox D. F. and Reid N. (2002). The Theory of the design of Experiments; Boca Raton, Chapman and Hall/CFC.
Cox. D.R. (1997). Introduction toYates and Cochran (1938) on the Analysis of Groups of Experiments:Breakthroughs in Statistics pp59-87.
Douglas C. M., (2000). Design and Analysis of Experiments; New York, John Wiley and sons, Inc.
Douglas K. N. Croy C. D., (2005). Methods for Addressing Missing Data in Psychiatric and Developmental Research; Journal of the American Academy of Child and Adolescent Psychiatry, 44[12], 1230-1240.
Fisher R. A., (1973). Statistical Methods and Scientific Inference; 3rd Edition New York Hafnew press.
Jarett, R. G. (1978). The Analysis of Designed Experiments with Missing Observations; Journal of the Royal Statistical Society App. Statist, 27(1) 38-46.
John.W.T., Taylor and Francis, (1984). The collected works of John W. Tukey, [1]: Journal of Mathematical Statistics.
Klaus. H. and Oscar K., (2007). Design and Analysis of experiments. Introduction to Experimental Design 2; New York John Wiley and sons, Inc.
Kutner M.H., Nachtsheim C., Neter J. and Li W., (2005). Applied Linear Statistical Models; New York, NY, McGrow-Hill/Irwin, a Business Unit of the McGrow-Hill Companies, Inc, 1221 Avenue of the Americas.
Leila B.O., (2006). Yet Another Approach for Completing Missing Values. Concept Lattices and their application, fourth international conference, CLA 2006, Runis, Runisia.
Levene H., (1960). Robust Test for Equality of Variances, in Contribution to Probability and Statistics: Essays in Honor of Harold Hotelling: Stanford University Press.
Little R.J. A. and Donald B. R., (1989), The Analysis of Social Science Data with Missing Values; Journal of Sociological Methods and Research, 18[2] [3], 292-326.
Sarhan A.E. and Greenberg B.G., (1956). Estimation of Location and Scale Parameters by Ordered Statistics from Single and Double Censored Samples, Part 1. Ann Math, Statist, [27], 427-51.
Shadish W.R., Cook HP. and Campebell D.T., (2002), Experimental and Quasi-Experimental Designs for Generalized Causal Inference; Boston New York, Houghton Mifflin Company.
Shapiro S.S., and Wilk M.B., (1965). An Analysis of Variance Test for Normality (Complete Samples), Biometrika [52], 591-611.