



# Estimation of Missing Data Using Convoluted Weighted Method in Nigeria Household Survey

Faweya Olanrewaju<sup>1,\*</sup>, Amahia Godwin Nwanzu<sup>2</sup>, Adeniran Adefemi Tajudeen<sup>2</sup>

<sup>1</sup>Department of Statistics, Ekiti State University, Ado Ekiti, Nigeria

<sup>2</sup>Department of Statistics, University of Ibadan, Ibadan, Nigeria

## Email address:

lanrefaweya@gmail.com (F. Olanrewaju)

\*Corresponding author

## To cite this article:

Faweya Olanrewaju, Amahia Godwin Nwanzu, Adeniran Adefemi Tajudeen. Estimation of Missing Data Using Convoluted Weighted Method in Nigeria Household Survey. *Science Journal of Applied Mathematics and Statistics*. Vol. 5, No. 2, 2017, pp. 70-77.

doi: 10.11648/j.sjams.20170502.12

**Received:** October 4, 2016; **Accepted:** October 12, 2016; **Published:** March 10, 2017

---

**Abstract:** The analysis of survey data becomes difficult in the presence of missing data. By the use of Least Squares and Stein Rule method, estimator for the parameters of interest can be obtained. In this study, proposed convoluted Weighted Least Squares and Stein Rule method is compared with some existing techniques where the data is considered missing completely at random (MCAR). The results show that other techniques are occasionally useful in estimating most of the parameter, but proposed (LSSR) technique perform better regardless of the percentage of the missing data under MCAR assumption.

**Keywords:** Missing Data, MCAR, Stein Rule, Least Squares, Convoluted Weighted Method

---

## 1. Introduction

Missing data problem is an inherent feature of all surveys and one of the greatest threats compromising the precision of most surveys estimate during design and analysis. It can impair the quality of survey statistics by threatening the ability to draw valid inference from the sample to the target population of the survey. The problem of missing data occurs when some or all of the responses are not collected for a sampled element or when some responses are deleted because they failed to satisfy edit constraints. It is common practice to distinguish between unit (or total) non response, when none of the survey responses are available for a sampled element and item non response, when some but not all of the responses are available. Total non response arises because of refusals, inability to participate, not at home, units closed, away on vacation, unit vacant or demolished and untraced units. Item non response arises because of item refusals, “don’t know” omissions and answers deleted in editing.

Over the years, attempts with varying degrees of success have been made in the literature to solve problem of missing data. The success of a particular technique is dependent on

the complexity of the problem and no technique is robust for all purposes of estimation but techniques are used indiscriminately.

This paper presents a robust technique of handling missingness and compare how well this technique performs with some existing ones in terms of what happens to mean, variance, correlation coefficient, skewness, and kurtosis under MCAR with different amount of missing data concerning Nigeria household survey.

## 2. Missing Data Mechanisms

### 2.1. Missing Completely at Random (MCAR)

The distribution of missing value R is assumed to be independent of both the target variable Y and auxiliary variable X. Thus

$$P(R/Y, X) = P(R) \quad (1)$$

### 2.2. Missing at Random (MAR)

In general, MAR occur when there is no direct relation between the target variable Y and response behavior R and the same time there is a relation between the auxiliary

variable and the response behavior R.

This is expressed as:

$$P(R/Y, X) = (R/Y^0, X) \tag{2}$$

### 2.3. Missing Not at Random (MNAR)

Missing data Mechanism where values are assumed to be related to the unobserved dependent variable vector  $Y_t^m$ , in addition to the remaining observed values is called Missing not at Random (MNAR). This is expressed as:

$$P(R/Y, X) = p(R/Y^m, Y^0, X) \tag{3}$$

## 3. Methodology: Techniques for Handling Missing Values

### 3.1. Least Square (Yates) Procedure

Yates (1933) proposed a technique that first estimating the parameters of the model with the help of the complete observations alone and obtaining the predicted values for the missing observations, when linearity and unbiasedness criteria of estimates are of interest [18]. The predicted value of the study variable is given as:

$$\hat{y}_{LS}^* = x_* b_c \tag{4}$$

where

$$b_c = (x_c' x_c)^{-1} x_c' y_{obs} \tag{5}$$

$x_*$  denote the k explanatory variable corresponding to the unobserved responses  $y_{mis}$

The variance of Least Squares (Yates) procedure is computed as follows:

$$\text{Var}(\hat{y}_{LS}) = \hat{\sigma}_{LS}^2 = \frac{1}{T-K} \left\{ \sum_{t=1}^{t_c} (y_t - \hat{y}_{LS}^*)^2 + \sum_{t=t_c+1}^T (\hat{y}_{LS}^* - \hat{y}_{LS}^*)^2 \right\} \tag{6}$$

$$\text{Var}(\hat{y}_{LS}) = \hat{\sigma}_{LS}^2 = \frac{\sum_{t=1}^{t_c} (y_t - \hat{y}_{LS}^*)^2}{(T-K)}$$

Which makes use of  $t_c$  observations but has T-K instead of  $t_c - k$  degrees of freedom, and  $\hat{y}_{LS}^*$  in the expression from (4)

### 3.2. Stein-Rule Strategy

This other method called Stein-Rule was proposed by James and Stein (1961), providing the following predictions:

$$\hat{y}_{SR}^* = \left(1 - \frac{KR_C}{(t_c - k + 2)b_c' x_c' x_c b_c}\right) x_* b_c \tag{7}$$

where:  $\hat{y}_c = x_c b_c$  and  $R_c = (y_c - x_c b_c)'(y_c - x_c b_c)$  is the residual sum of square and k is a positive non stochastic scalar. [18]

The variance of the Stein-Rule procedure is given as:

$$\text{Var}(\hat{y}_{SR}^*) = \sigma_{stein}^2 = \frac{\sum_{t=1}^{t_c} (y_t - \hat{y}_{SR}^*)^2}{(T-K)} \tag{8}$$

where:  $\hat{y}_{SR}^*$  is the expression from (7)

### 3.3. Proposed Convoluted Weighted Method

If  $\hat{Y}_1^* = \phi_1(x)$  and  $\hat{Y}_2^* = \phi_2(x)$  are two different functions (models) in estimating the missing values of the study variables.

Let us define our target model as:

$$\hat{Y} = \alpha_1 \phi_1(x) + \alpha_2 \phi_2(x) \tag{9}$$

Which is a linear combination of two existing models.

Where  $\alpha_1 + \alpha_2 = 1$  and  $\alpha_1$  is a non stochastic scalar between 0 and 1; see [18]. The value of  $\alpha_1$  may reflect the weight been given to the prediction of first model value in relation to the prediction of second model values.

This implies that,  $\alpha_2 = 1 - \alpha_1$

Also let  $P = \hat{Y}^2 = [\alpha_1 \phi_1(x) + \alpha_2 \phi_2(x)]^2$  be a quadratic combination of the models, we have

$$P = \hat{Y}_*^2 = \alpha_1^2 (\phi_1(x))^2 + \alpha_2^2 (\phi_2(x))^2 + 2\alpha_1 \alpha_2 \phi_1(x) \phi_2(x) \tag{10}$$

$$\text{If } \hat{Y}_1^* = \hat{Y}_{LS}^* = x_* b_c \tag{11}$$

$$\hat{Y}_2^* = \hat{Y}_{SR}^* = \left(1 - \frac{KR_C}{(t_c - k + 2)b_c' x_c' x_c b_c}\right) x_* b_c \tag{12}$$

where:

$b_c = (x_c' x_c)^{-1} x_c' y_{obs}$  (Least Squares Method).

$R_c = (y_c - x_c b_c)'(y_c - x_c b_c)$  (Stein Rule Model).

$\alpha_2 = 1 - \alpha_1$

$t_c$  = Number of observed cases

k = Number of explanatory variables (which is a positive non stochastic scalar)

Then,

$$P = \alpha_1^2 [x_* b_c]^2 + \alpha_2^2 \left[1 - \frac{KR_C}{(t_c - k + 2)b_c' x_c' x_c b_c}\right]^2 [x_* b_c]^2 + (2\alpha_1 - 2\alpha_1^2) [x_* b_c] \left[1 - \frac{KR_C}{(t_c - k + 2)b_c' x_c' x_c b_c}\right] x_* b_c \tag{13}$$

Taking the partial derivative of the expression (13) above with respect to parameter  $\alpha_1$ , we have

$$\frac{dP}{d\alpha_1} = 2\alpha_1 [x_* b_c]^2 - 2[(1 - w)x_* b_c]^2 + 2\alpha_1 [(1 - w)x_* b_c]^2 + 2([x_* b_c]) ((1w)x_* b_c) - 4\alpha_1 [x_* b_c] [(1 - w)x_* b_c] \tag{14}$$

where:

$$W = \frac{KR_C}{(t_c - k + 2)b_c' x_c' x_c b_c}$$

At turning point  $\frac{d_p}{d\alpha_1} = 0$ , therefore, setting (14) to zero, we have

$$2[(1-w)x_*b_c]^2 + 2(x_*b_c)((1-w)x_*b_c) = 0 \tag{15}$$

Therefore, Solving for  $\alpha_1$  in (15), we have

$$2\alpha_1[x_*b_c]^2 + 2\alpha_1[(1-w)x_*b_c]^2 - 4\alpha_1[x_*b_c][(1-w)x_*b_c] -$$

$$\alpha_1 = \frac{((1-w)x_*b_c)^2 - (x_*b_c)((1-w)x_*b_c)}{[x_*b_c]^2 + ((1-w)x_*b_c)^2 - 2[x_*b_c][(1-w)x_*b_c]}$$

$$\text{Hence, } \hat{\alpha}_1 = \frac{\left(1 - \frac{KR_C}{(t_c - k + 2)b_c^1 x_c^1 b_c}\right) x_* b_c - (x_* b_c) \left(1 - \frac{KR_C}{(t_c - k + 2)b_c^1 x_c^1 b_c}\right) x_* b_c}{[x_* b_c]^2 + \left(1 - \frac{KR_C}{(t_c - k + 2)b_c^1 x_c^1 b_c}\right) x_* b_c - 2[x_* b_c] \left(1 - \frac{KR_C}{(t_c - k + 2)b_c^1 x_c^1 b_c}\right) x_* b_c}$$

$$\hat{\alpha}_1 = \frac{\left(1 - \frac{KR_C}{(t_c - k - 2)b_c^1 x_c^1 b_c}\right) (x_* b_c)^2 - \left(1 - \frac{KR_C}{(t_c - k - 2)b_c^1 x_c^1 b_c}\right) (x_* b_c)^2}{\left[\left(1 - \frac{KR_C}{(t_c - k - 2)b_c^1 x_c^1 b_c}\right) x_* b_c - x_* b_c\right]^2} \tag{16}$$

The predicted values of the study variable using the proposed model is given as:

$$\hat{Y}_{LSSR}^* = \hat{\alpha}_1 \hat{Y}_{LS}^* + \hat{\alpha}_2 \hat{Y}_{SR}^* \tag{17}$$

where:

$\hat{Y}_{LS}^*$ ,  $\hat{Y}_{SR}^*$  is as shown in (4), (7) respectively and

$\hat{\alpha}_2 = 1 - \hat{\alpha}_1$

$t_c$  = Indicates the number of observed cases

$K$  = Number of explanatory variables (which is a positive non stochastic scalar)

Thus, the proposed weighted convoluted model is:

$$\hat{Y}_{LSSR}^* = \hat{\alpha}_1 \hat{Y}_{LS}^* + (1 - \hat{\alpha}_1) \hat{Y}_{SR}^* \tag{18}$$

### 4. Efficiency Comparison

If the data are complete, then  $S^2 = \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T-K}$  is the corresponding estimator of variance ( $\sigma^2$ ). If  $T-t_c$  cases are incomplete, that is, observation  $y_{mis}$  are missing in the model, then the variance  $\sigma^2$  can be estimated using the complete case estimator as:

$$\hat{\sigma}_c^2 = \frac{\sum_{t=1}^{t_c} (y_t - \hat{y}_t)^2}{t_c - K} \tag{19}$$

#### 4.1. Using Least Square Method

If the missing data are imputed using Least Square (Yates) method, then we have the estimator

$$\begin{aligned} \hat{\sigma}_{LS}^2 &= \frac{1}{(T-K)} \left\{ \sum_{t=1}^{t_c} (y_t - \hat{y}_{LS})^2 + \sum_{t=1}^T (y_{LS} - \hat{y}_{LS})^2 \right\} \tag{20} \\ &= \frac{\sum_{t=1}^{t_c} (y_t - \hat{y}_{LS})^2}{T-K} \\ &= \frac{\sum_{t=1}^{t_c} (y_t - x_* b_c)^2}{T-K} \tag{21} \end{aligned}$$

which makes use of  $t_c$  observations but has  $T-K$  instead of  $t_c-k$  degrees of freedom. As

$$\hat{\sigma}_{LS}^2 = \hat{\sigma}_c^2 \frac{t_c - k}{T - K} < \hat{\sigma}_c^2 \tag{22}$$

#### 4.2. Using Stein Rule Method

If the missing data are imputed using Stein Rule approach, then we have the estimate

$$\hat{\sigma}_{SR}^2 = \frac{\sum_{t=1}^{t_c} (y_t - (1 - \frac{KR_C}{t_c - k + 2}) x_* b_c)^2}{T - K} \tag{23}$$

#### 4.3. Efficiency Comparison: Least Squares Versus Stein Rule Techniques

The following three possible conditions will hold iff  $x_* b_c > 0$ :

(i) when  $\frac{KR_C}{t_c - k + 2} > 0$ , then  $\hat{\sigma}_{SR}^2 > \hat{\sigma}_{LS}^2$  Hence, Least Squares technique will be more efficient

(ii) when  $\frac{KR_C}{t_c - k + 2} = 0$ , then  $\hat{\sigma}_{SR}^2 = \hat{\sigma}_{LS}^2$  Hence, both techniques perform at the same level.

(iv) when  $\frac{KR_C}{t_c - k + 2} < 0$ ,  $\hat{\sigma}_{SR}^2 < \hat{\sigma}_{LS}^2$  Hence, Stein Rule method will be more efficient.

#### 4.4. Efficiency Comparison: Proposed Technique Versus Least Squares and Stein Rules Techniques

If the missing data are imputed using the proposed weighted least square stein rule (LSSR) method, then we have the estimator

$$\hat{\sigma}_{LSSR}^2 = \frac{\sum_{t=1}^{t_c} \{y_t - [\hat{\alpha}_1 x_* b_c + (1 - \hat{\alpha}_1) (1 - \frac{KR_C}{t_c - k + 2b_c^1 x_c^1 b_c}) x_* b_c]\}^2}{T - K} \tag{24}$$

Comparison of variance Estimates from Least Square, Stein Rule and Proposed Technique.

(i) When  $\hat{\alpha}_1 \rightarrow 1$ , then  $\sigma_{LSSR(Proposed)}^2 \rightarrow \sigma_{LS(Least Square)}^2$

(ii) When  $\hat{\alpha}_1 \rightarrow 0$ , then  $\sigma_{LSSR(Proposed)}^2 \rightarrow \sigma_{SR(Stein Rule)}^2$

### 5. Performance Criteria for the Techniques

The criteria comprises of the following:

(i) Descriptive and analytic statistics obtained under

MCAR with different percentage of missing data.

(ii) Bias of the parameter estimate from each technique such as;

$$\begin{aligned} \text{BIAS}_{SR}(\hat{\mu}) &= \hat{\mu}_{i_{SR}} - \mu \\ \text{BIAS}_{LS}(\hat{\mu}) &= \hat{\mu}_{i_{LS}} - \mu \\ \text{BIAS}_{LSSR}(\hat{\mu}) &= \hat{\mu}_{i_{LSSR}} - \mu \end{aligned}$$

where  $\hat{\mu}_{i_{SR}}$  = Estimated mean at i% of data missing using stein Rule method. i= 5%, 12%, 23% and 44%.  $\mu$ =Actual mean value when data is complete.

(iii) The Root Mean Square Error for each technique as:

$$\text{RMSE}_{SR(\text{SteinRule})} = [\text{Bias}^2_{SR(\text{Stein Rule})} + \text{Var}_{SR(\text{Stein Rule})}]^{\frac{1}{2}}$$

$$\text{RMSE}_{LS(\text{LeastSquare})} = [\text{Bias}^2_{LS(\text{Least Square})} + \text{Var}_{LS(\text{Least Square})}]^{\frac{1}{2}}$$

$$\text{RMSE}_{LSSR(\text{proposed})} = [\text{Bias}^2_{LSSR(\text{proposed})} + \text{Var}_{LSSR(\text{proposed})}]^{\frac{1}{2}}$$

The technique with minimum (RMSE) is adjudged the best.

## 6. Numerical Illustration for the Proposed Technique

A simple random sample of n = 100 households was selected from the records of survey data on “household income” from Akure North Local Government, Iju/Ita-Ogbolu in Ondo state, Nigeria to evaluate the performance of the proposed model with some existing techniques of handling missing data under MCAR using different percentage of missing data.

Three demographic variables; Y (income N’000), Age ( $X_2$ ) and year of schooling ( $X_1$ ) were considered. The Y variable was generated as a combination of explanatory variables with added random components. Then, differing amounts were deleted at random causing MCAR data which had 0,5,12,23 and 44% missing data.

## 7. Results and Discussion

**Table 1.** Performance of Some Missing Data Techniques for Parameter Estimates when Differing Amount of Data are Missing Under MCAR Assumption of Missingness.

| ESTIMATED PARAMETERS          | PERCENTAGE OF MISSINGNESS | MISSING DATA TECHNIQUES |                   |                 |                 |                |
|-------------------------------|---------------------------|-------------------------|-------------------|-----------------|-----------------|----------------|
|                               |                           | MI (Mean Imputation)    | LS (Least Square) | SR (Stein Rule) | LSSR (Proposed) | LW (List Wise) |
| MEAN ( $\bar{y}$ )            | 0%                        | 13.814                  | 13.814            | 13.814          | 13.814          | 13.814         |
|                               | 5%                        | 13.6198                 | 13.79885          | 13.804          | 13.80323        | 13.68989       |
|                               | 12%                       | 13.6044                 | 13.8055           | 13.41354        | 13.80565        | 13.68787       |
|                               | 23%                       | 14.1119                 | 13.79732          | 13.79741        | 13.8074         | 14.30282       |
|                               | 44%                       | 13.7306                 | 13.80303          | 13.87161        | 13.74924        | 14.11842       |
| CORRELATION ( $\rho$ )        | 0%                        | 0.946                   | 0.946             | 0.946           | 0.946           | 0.946          |
|                               | 5%                        | 0.881                   | 0.967             | 0.967           | 0.967           | 0.978          |
|                               | 12%                       | 0.686                   | 0.966             | 0.793           | 0.966           | 0.978          |
|                               | 23%                       | 0.596                   | 0.981             | 0.981           | 0.98            | 0.976          |
|                               | 44%                       | 0.488                   | 0.518             | 0.886           | 0.961           | 0.994          |
| VARIANCE ( $\hat{\sigma}^2$ ) | 0%                        | 46.577                  | 46.577            | 46.577          | 46.577          | 46.577         |
|                               | 5%                        | 45.94065                | 46.63859          | 46.67568        | 46.67109        | 48.14344       |
|                               | 12%                       | 38.10063                | 47.8907           | 41.78268        | 47.88495        | 42.64912       |
|                               | 23%                       | 35.14046                | 48.80046          | 48.80107        | 48.84908        | 44.57015       |
|                               | 44%                       | 28.59472                | 49.14666          | 48.6712         | 48.83988        | 48.94685       |
| SKEWNESS ( $S_k$ )            | 0%                        | 0.217                   | 0.217             | 0.217           | 0.217           | 0.217          |
|                               | 5%                        | 0.291648                | 0.222904          | 0.221543        | 0.221757        | 0.258074       |
|                               | 12%                       | 0.190889                | 0.248072          | 0.124799        | 0.248259        | 0.144436       |
|                               | 23%                       | 0.169918                | 0.227823          | 0.227841        | 0.223639        | 0.071633       |
|                               | 44%                       | 0.194166                | 0.259757          | 0.243059        | 0.282537        | -0.00069       |
| KURTOSIS (K)                  | 0%                        | 2.616                   | 2.616             | 2.616           | 2.616           | 2.616          |
|                               | 5%                        | 2.704214                | 2.611784          | 2.608406        | 2.608843        | 2.580429       |
|                               | 12%                       | 2.793288                | 2.653841          | 2.567728        | 2.653645        | 2.496646       |
|                               | 23%                       | 3.089743                | 2.608665          | 2.608734        | 2.602258        | 2.45105        |
|                               | 44%                       | 3.995174                | 2.626688          | 2.660419        | 2.681545        | 2.387055       |
| COEFFICIENT OF VARIATION (CV) | 0%                        | 49.62                   | 49.62             | 49.62           | 49.62           | 49.62          |
|                               | 5%                        | 49.76544                | 49.4914           | 49.4926         | 49.49294        | 50.68371       |
|                               | 12%                       | 45.37187                | 50.12722          | 48.18977        | 50.12365        | 47.71109       |
|                               | 23%                       | 42.00667                | 50.63109          | 50.63107        | 50.61933        | 46.67672       |
|                               | 44%                       | 38.94515                | 50.78933          | 50.29317        | 50.82865        | 49.55372       |



Figure 1. Graph of mean for MCAR imputed data by Amount of data missing.

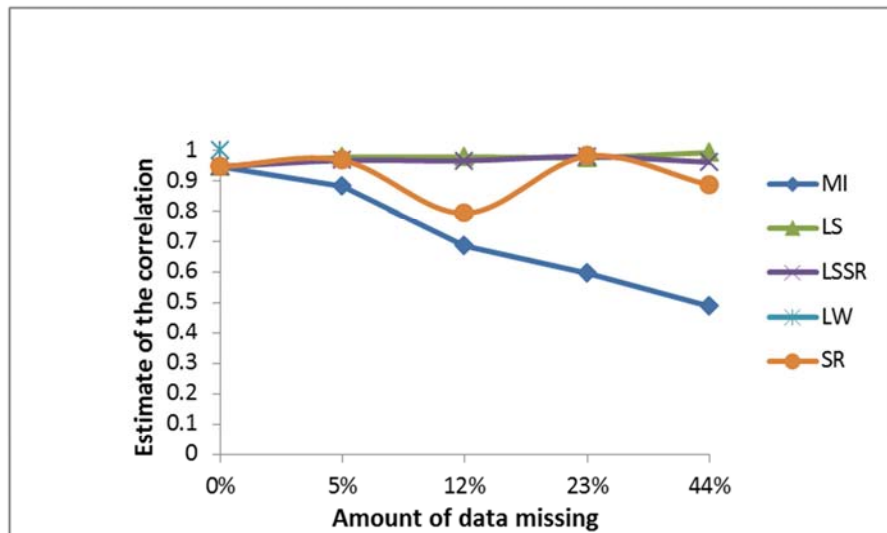


Figure 2. Graph of correlation for MCAR imputed data by Amount of data missing.

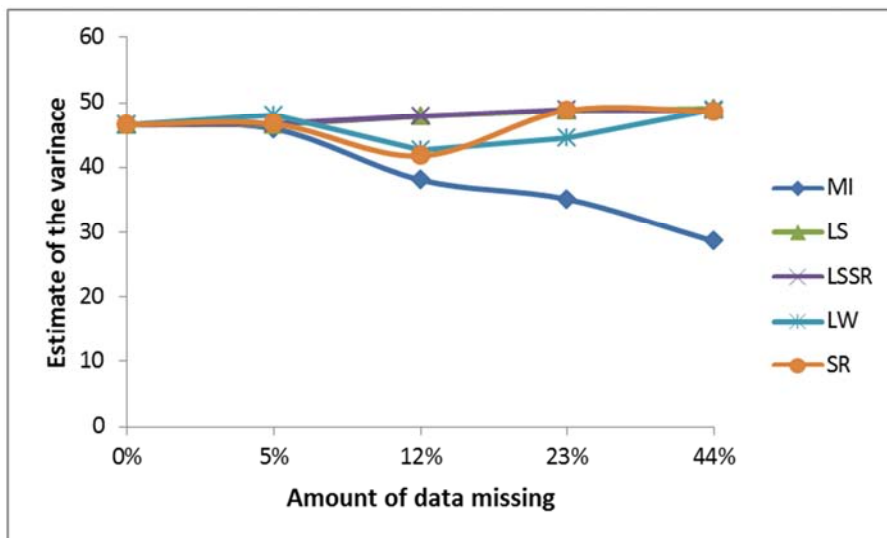


Figure 3. Graph of variance for MCAR imputed data by Amount of data missing.



Figure 4. Graph of skewness for MCAR imputed data by Amount of data missing.

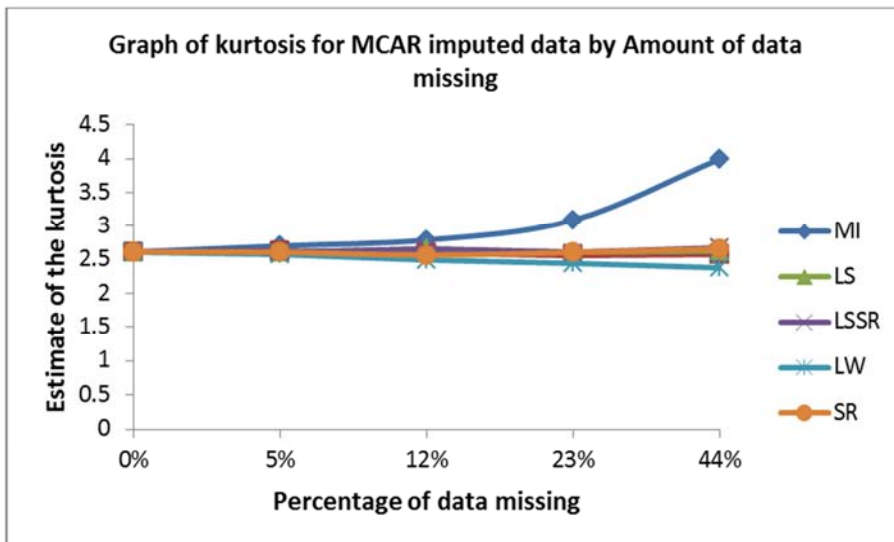


Figure 5. Graph of kurtosis for MCAR imputed data by Amount of data missing.

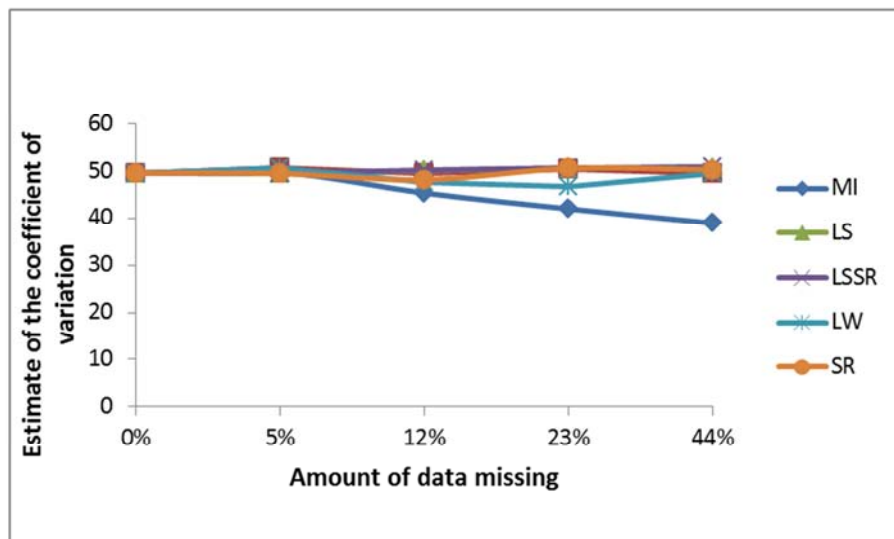
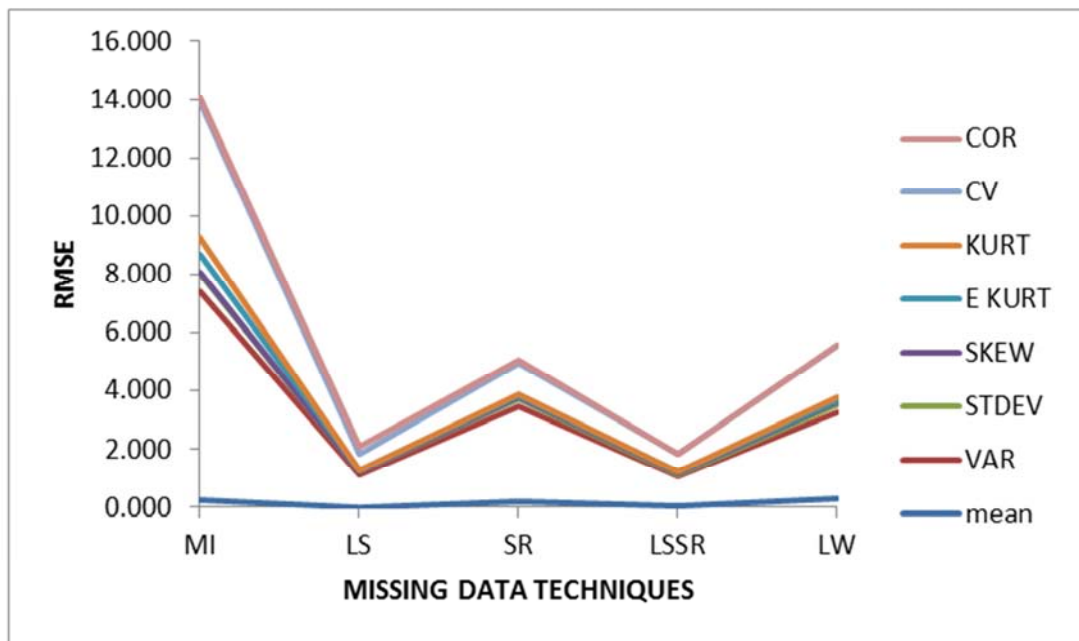


Figure 6. Graph of coefficient of variation for MCAR imputed data by Amount of data missing.

**Table 2.** Performance of Missing Techniques for Parameter Estimates Under Missing Completely at Random (MCAR).

| Average Parameter Estimate (with RMSE in Parenthesis) |                      |                   |                   |                   |                   |
|---|----------------------|-------------------|-------------------|-------------------|-------------------|
| PARAMETER   | MI (Mean Imputation) | LS (Least Square) | SR (Stein Rule)   | LSSR (Proposed)   | LW (List Wise)    |
| mean = 13.814   | 13.767<br>(0.237)    | 13.801<br>(0.004) | 13.722<br>(0.208) | 13.791<br>(0.028) | 13.950<br>(0.310) |
| VAR = 46.577  | 36.944<br>(7.194)    | 48.119<br>(1.120) | 46.483<br>(3.281) | 48.061<br>(1.031) | 46.077<br>(2.974) |
| STDEV = 6.8247  | 6.056<br>(0.593)     | 6.936<br>(0.081)  | 6.815<br>(0.244)  | 6.932<br>(0.075)  | 6.785<br>(0.220)  |
| SKEW = 0.217  | 0.212<br>(0.054)     | 0.240<br>(0.017)  | 0.204<br>(0.054)  | 0.244<br>(0.028)  | 0.118<br>(0.110)  |
| KURT = 2.616  | 3.146<br>(0.590)     | 2.625<br>(0.021)  | 2.611<br>(0.038)  | 2.637<br>(0.038)  | 2.479<br>(0.081)  |
| COV = 49.62   | 44.022<br>(4.642)    | 50.260<br>(0.585) | 49.652<br>(1.085) | 50.266<br>(0.594) | 48.656<br>(1.801) |
| COR = 0.946   | 0.663<br>(0.167)     | 0.858<br>(0.227)  | 0.907<br>(0.087)  | 0.969<br>(0.008)  | 0.982<br>(0.008)  |



**Figure 7.** Graph of Root Mean Square Error of some missing data techniques under MCAR missingness assumption.

**Table 3.** Summary of the Result from the Figure 1-6 of Performance of some Missing Data Techniques under MCAR as Percentage of Missing Value Increases.

| Estimated parameter | MI (Mean Imputation)                          | LS (Least Square)                            | SR (Stein Rule)                               | LSSR (Proposed)                              | LW (List Wise)                                |
|---------------------|---|--|---|--|---|
| Mean                | High percentage discrepancy in the true value | Approximately constant (within target value) | Approximately constant (within target value)  | Approximately constant (within target value) | High percentage discrepancy in the true value |
| Correlation         | High percentage discrepancy in the true value | Approximately constant (within target value) | Some percentage discrepancy of true value     | Approximately constant (within target value) | High percentage discrepancy in the true value |
| Variance            | High percentage discrepancy in the true value | Approximately constant (within target value) | High percentage discrepancy in the true value | Approximately constant (within target value) | Little percentage of discrepancy              |
| Skewness            | High percentage discrepancy in the true value | Approximately constant (within target value) | High percentage discrepancy in the true value | Approximately constant (within target value) | High percentage discrepancy in the true value |
| Kurtosis            | High percentage discrepancy in the true value | Approximately constant (within target value) | Approximately constant (within target value)  | Approximately constant (within target value) | High percentage discrepancy in the true value |
| CV                  | High percentage discrepancy in the true value | Approximately constant (within target value) | Approximately constant (within target value)  | Approximately constant (within target value) | Little percentage of discrepancy              |

Remark: Although, other techniques are occasionally useful in estimating most of the true parameter, proposed (LSSR) technique perform better regardless of the percentage of the missing data under MCAR assumption considered the results from the criteria for selection. However, for mean imputation (MI) and list wise (LW), there is higher percentage of discrepancy in the true values of most of the parameters. Hence, the proposed technique preserve most of the parameters structure within the data. That is there is almost no change in the mean, variance, skewness, kurtosis, coefficient of variation and correlation coefficient under MCAR assumption using the proposed model of imputation.

## 8. Conclusion

Although other procedures are occasionally useful, proposed (LSSR) technique performed better regardless of the percentage of the existing data under MCAR nature of missingness.

Considered the result from the criteria for selection, the proposed model reduces the variability around the true parameter value without discarding the linearity and unbiasedness criteria.

---

## References

- [1] Amahia, G. N (2010) Factors, Prevention and Correction Methods for Non-Response in Sample Surveys. Central Bank of Nigeria, Journal of Applied Statistics 1 (1), pp 79-89, (Nigeria).
- [2] Arbuckle, J. L. (1996). Full Information Likelihood Estimation in the Presence of Incomplete Data.
- [3] Beale, E. M. L., & Little, R. J. A (1975). Missing Values in Multivariate Analysis. Journal of the Royal Statistical Society, Series B, 37, 129-146.
- [4] Cochran, W. (1968). "The Effective of Adjustment by Sub Classification in Removing Bias in Observational Studies." Biometrics, 24, pp. 295-313.
- [5] Cool, A. L. (2000). A Review of Methods for Dealing with Missing Data. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Dallas, TX. (Eric Document Reproduction Service No. ED 438 311).
- [6] Diggle P. J Testing for Random Dropouts in Repeated Measurement Data. Biometrics, 45 (1989) 1255-1258.
- [7] Diggle P. J., Heagerty P, Liang, K. Y. et al. Analysis of Longitudinal Data. 2nd ed. Oxford University Press Inc., (2002).
- [8] Dixon W. J BMDP Statistical software. Los Angeles. University of California Press, (1988).
- [9] Ender C. K., and Bandalos, D. L (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. Structural Equation Modeling, 8(3), 430-457.
- [10] Groves, R Couper M. (1998); Non-response Interview Survey, New York Wiley.
- [11] Groves, R Singer. and Corning, A (2000); Leverage- Saliency Theory of Survey Participation Description and an Illustration. Public Opinion Quarterly, 64, 299-308.
- [12] Groves, R. Cialdini R and Couper M (1992): "Understanding the Decision to participate in a Survey" The public Opinion Quarterly, 54 (4), 475-495.
- [13] Howell, D. C (2007) The Analysis of Missing Data. Handbook of social science Methodology London: Sage. Return.
- [14] Kalton, G. and Flores-Cervantes, I (2003), "Weighting Methods" Journal of Official statistics, 19, pp. 81-97.
- [15] Kim, J. O., & Curry, J. (1977). The Instrument of Missing Data in Multivariate Analysis Sociological Methods & Research, 6 (2), 215-240.
- [16] Little R. J. A. A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of American Statistical Association, 83 (1988), 1198-2002.
- [17] Madlow, W. G., Nisselson, H., Olkin, I. (Eds.), 1983. Incomplete Data in Sample Surveys. Report and Case Studies, vol.1. Academic Press, New York.
- [18] Rao and Toutenburg. H (1999) Linear Models: Least Square and Alternatives. New York: Springer-Verlag.
- [19] Quinten, A., Raaijmakers, W., 1999. Effectiveness of Different Missing Data Treatments in Survey with Likert-Type Data: Introducing the Relative Mean Substitution Approach. Educational and Psychological Measurement 59 (5), 725-748.
- [20] Schafer J. L., Graham J. W. Missing Data: Over View of the State of the Art. Psychological Methods, 7 (2002), 147: 177.
- [21] Schafer, J (2002). Dealing with Missing Data. Research Letter Information Mathematics Science. Vol. 3, pp 153-160.
- [22] Tshering, S., Okazaki T and Endo S., March 2013: A Method to Identify Missing Data Mechanism in Incomplete Dataset. IJCSNS International Journal of Computer Science and Network Security, Vol. 13 No14 Page (14-21).
- [23] Utazi C. E., Onyeagu S. I. & Osuji G. A (2010). On the Efficiency of Some Techniques for Estimating Covariance and Correlation Matrices from Incomplete Data. Journal of the Nigerian Statistical Association Vol. 22, 44-63.