MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 3, pp. 872–881 (2025)

Study of some exponential-inverse sine-logarithmic imputation techniques under missing data using simulated data

Singh B. K., Das J.

Mathematics Department,
North Eastern Regional Institute of Science and Technology,
Nirjuli, Arunachal Pradesh 791109, India

(Received 10 February 2025; Revised 11 September 2025; Accepted 13 September 2025)

Non-response in a survey refers to the absence of data from selected participants who fail to provide information for various reasons. Imputation is one of the most effective techniques to address non-response in sample surveys and ensure data completeness. The most commonly used imputation methods are mean imputation, ratio imputation, and the compromised imputation method. Mean imputation replaces all missing values with the mean of the responded values, thus reducing the variability in the data set. To maintain a proportional link between variables, the ratio imputation technique is useful, although it assumes a strong linear relationship between the study and auxiliary variables, which may not always hold. If violated, this can lead to biased results. The compromised imputation method combines several techniques but still has limitations and may produce biased outcomes when underlying assumptions are not met. To address these issues, we propose three Exponential-Inverse Sine-Logarithmic (ESL) imputation techniques along with their corresponding point estimators. We derive the bias and mean square error (MSE) of the proposed estimators and evaluate their performance both theoretically and numerically in comparison with existing methods. Additionally, simulated population data sets were generated using statistical software to conduct simulation studies. Percentage relative efficiencies (PRE) were calculated to compare the performance of all estimators with respect to the mean and ratio methods. Based on the results, we conclude that the proposed imputation techniques outperform the existing ones.

Keywords: *simple random sampling; imputation; bias; mean square error (MSE); percentage relative efficiency (PRE); simulation study.*

2010 MSC: 62D05, 94A20 **DOI:** 10.23939/mmc2025.03.872

1. Introduction

In sample surveys, missing data or missing values occur when a chosen respondent refuses to participate in a survey for various reasons. Various techniques are available to handle the problems of missing data. One of the best techniques to deal with missing data is imputation. In finite population sampling, Meeden [1] discussed a theoretic approach to imputation. Lee and Sardal proposed an imputation technique known as the ratio method of imputation. Later, an improved imputation technique named as the compromised imputation technique was developed in [2]. Singh et al. [3] proposed exponential type imputation technique for missing observations. Prasad [4] proposed product exponential method of imputation. Singh et al. [5] developed some logarithmic and sine-type imputation techniques. Pandey et al. [6] proposed some new logarithmic-type imputation methods for handling missing data. Later, Singh and Gogoi [7] proposed exponential dual to ratio type compromised imputation techniques.

Consider a sample S of size n drawn from a finite population Ω of size N without replacement. Let Y be the study variable and X be the auxiliary variable. $\overline{Y}$ and $\overline{X}$ are the population means of the variables Y and X respectively given by

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Let the sample consist of r responding units, which form the set Re, and $(n-r)$ non-responding units, which belong to the set Re^C.

2. Some known methods of imputation

2.1. Mean method

In this imputation method, each non-response data is replaced with the mean of the responded data. Here

$$y_{.i} = \begin{cases} y_i, & \text{if } i \in \mathrm{Re}, \\ \overline{y}_r, & \text{if } i \in \mathrm{Re}^c. \end{cases}$$

The unit estimator of the above imputation method is given by

$$\overline{y}_s = \frac{1}{n} \sum_{i \in S} y_{.i} = \overline{y}_r,$$

where $\overline{y}_r = \frac{1}{r} \sum_{i \in \mathrm{Re}} y_{.i}$.

Bias, $B(\cdot)$ of the above point estimator y_r is

$$B(\overline{y}_r) = 0.$$

The MSE of the above unit estimator y_r is

$$\mathrm{MSE}\,(\overline{y}_r) = \overline{Y}^2 \lambda_{r,N} C_Y^2.$$

2.2. Ratio method

This method, proposed by Lee and Sardal [8],

$$y_{.i} = \begin{cases} y_i, & \text{if } i \in \mathrm{Re}, \\ \hat{a} x_i, & \text{if } i \in \mathrm{Re}^c, \end{cases}$$

where $\hat{a} = \frac{\sum_{i \in \mathrm{Re}} y_i}{\sum_{i \in \mathrm{Re}} x_i}$.

The unit estimator of the above imputation method is given by

$$\overline{y}_{\mathrm{RAT}} = \overline{y}_r \frac{\overline{x}_n}{\overline{x}_r},$$

where $\overline{y}_r = \frac{1}{r} \sum_{i \in \mathrm{Re}} y_{.i}$, $\overline{x}_r = \frac{1}{r} \sum_{i \in \mathrm{Re}} x_i$, $\overline{x}_n = \frac{1}{n} \sum_{i \in S} x_i$.

The bias of the above unit estimator y_{RAT} is

$$B\left(\overline{y}_{\mathrm{RAT}}\right) = \overline{Y}^2 \lambda_{r,n} \left(1 - \rho_{YX}\right) C_X^2.$$

The MSE of the above unit estimator y_{RAT} is

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{RAT}}\right) = \overline{Y}^2 \left\{ \lambda_{r,N} C_Y^2 + \lambda_{r,n} \left(1 - 2\phi_{YX}\right) C_X^2 \right\}.$$

2.3. Compromised method

Under this imputation technique, the data take the form

$$y_{.i} = \begin{cases} \alpha \frac{n}{r} y_i + (1-\alpha) \hat{a} x_i, & \text{if } i \in \mathrm{Re}, \\ (1-\alpha) \hat{a} x_i, & \text{if } i \in \mathrm{Re}^c, \end{cases}$$

where α is a chosen constant.

The unit estimator of the above imputation method is given by

$$\overline{y}_{\mathrm{COMP}} = \alpha \overline{y}_r + (1-\alpha) \overline{y}_r \frac{\overline{x}_n}{\overline{x}_r}.$$

The bias of the above point estimator y_{COMP} is

$$B\left(\overline{y}_{\mathrm{COMP}}\right) = \overline{Y} \lambda_{r,n} \left(1 - \alpha\right) \left(1 - \phi_{YX}\right) C_X^2.$$

The MSE $M(\cdot)$ of the above unit estimator y_{COMP} is

$$M\left(\overline{y}_{\mathrm{COMP}}\right) = \overline{Y}^2 \left[\lambda_{r,N} C_Y^2 + \lambda_{r,n} \left\{ (1-\alpha)^2 - 2(1-\alpha)\phi_{YX} \right\} C_X^2 \right].$$

The MSE of the resultant estimator achieves its minimum at $\alpha_{\mathrm{opt}} = 1 - \phi_{YX}$.

Hence, the optimum mean square error (MSE) is given by

$$\text{MSE}\left(\overline{y}_{\text{COMP}}\right)_{\text{opt}} = \overline{Y}^2\left(\lambda_{r,N} - \lambda_{r,n}\rho_{YX}^2\right)C_Y^2,$$

where ρ_{YX} is the Pearson's correlation coefficient between the study variable Y and the auxiliary variable X and

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2, \quad S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2$$

are the population mean squares of the study variable Y and the auxiliary variable X, respectively. Let

$$\phi_{YX} = \rho_{YX}\frac{C_Y}{C_X},$$

where $C_X^2 = \frac{S_X^2}{\overline{X}^2}$ and $C_Y^2 = \frac{S_Y^2}{\overline{Y}^2}$ and $a = \frac{n}{N-n}$.

Consider $\lambda_{n,N} = \left(\frac{1}{n} - \frac{1}{N}\right)$, $\lambda_{r,N} = \left(\frac{1}{r} - \frac{1}{N}\right)$, $\lambda_{r,n} = \left(\frac{1}{r} - \frac{1}{n}\right)$.

3. Proposed imputation techniques and corresponding estimators

The three imputation techniques with their corresponding point estimators are suggested as follows

$$\textbf{(1)} \quad (y_{.i})_I = \begin{cases} y_i, & \text{if } i \in \text{Re}, \\ \frac{1}{n-r}\overline{y}_r\left[n\exp\left[\alpha\sin^{-1}\left\{1 - \left(\frac{\ln\overline{X}}{\ln\overline{x}_r}\right)^{\beta}\right\}\right] - r\right], & \text{if } i \in \text{Re}^c, \end{cases}$$

where α and β are suitably chosen constants.

The point estimator of the population mean $\overline{Y}$ in the proposed imputation method is

$$\overline{y}_{\text{ESL1}} = \overline{y}_r\exp\left[\alpha\sin^{-1}\left\{1 - \left(\frac{\ln\overline{X}}{\ln\overline{x}_r}\right)^{\beta}\right\}\right], \tag{1}$$

$$\textbf{(2)} \quad (y_{.i})_{II} = \begin{cases} y_i, & \text{if } i \in \text{Re}, \\ \frac{1}{n-r}\overline{y}_r\left[n\exp\left[\alpha\sin^{-1}\left\{1 - \left(\frac{\ln\overline{X}}{\ln\overline{x}_n}\right)^{\beta}\right\}\right] - r\right], & \text{if } i \in \text{Re}^c, \end{cases}$$

where α and β are appropriately chosen constants.

The point estimator of the population mean $\overline{Y}$ in the proposed method of imputation is

$$\overline{y}_{\text{ESL2}} = \overline{y}_r\exp\left[\alpha\sin^{-1}\left\{1 - \left(\frac{\ln\overline{X}}{\ln\overline{x}_n}\right)^{\beta}\right\}\right], \tag{2}$$

$$\textbf{(3)} \quad (y_{.i})_{III} = \begin{cases} y_i, & \text{if } i \in \text{Re}, \\ \frac{1}{n-r}\overline{y}_r\left[n\exp\left[\alpha\sin^{-1}\left\{1 - \left(\frac{\ln\overline{x}_n}{\ln\overline{x}_r}\right)^{\beta}\right\}\right] - r\right], & \text{if } i \in \text{Re}^c, \end{cases}$$

where α and β are suitably chosen constants.

The point estimator of the population mean $\overline{Y}$ in the proposed method of imputation is

$$\overline{y}_{\text{ESL3}} = \overline{y}_r\exp\left[\alpha\sin^{-1}\left\{1 - \left(\frac{\ln\overline{x}_n}{\ln\overline{x}_r}\right)^{\beta}\right\}\right]. \tag{3}$$

4. Properties of the proposed point estimators

To determine the bias and mean square error (MSE) of the proposed imputed estimators, we write

$$e_1 = \frac{\overline{y}_r - \overline{Y}}{\overline{Y}}, \quad e_2 = \frac{\overline{x}_r - \overline{X}}{\overline{X}}, \quad e_3 = \frac{\overline{x}_n - \overline{X}}{\overline{X}}$$

such that $|e_i| < 1$, where $i = 1, 2, 3$.

Here, we have

$$E(e_1) = E(e_2) = E(e_3) = 0$$

and

$$E(e_1^2) = \lambda_{r,N}C_Y^2, \quad E(e_2^2) = \lambda_{r,N}C_X^2, \quad E(e_3^2) = \lambda_{n,N}C_X^2,$$

$$E(e_1e_2) = \lambda_{r,N}\rho_{YX}C_YC_X, \quad E(e_2e_3) = \lambda_{n,N}C_X^2, \quad E(e_1e_3) = \lambda_{n,N}\rho_{YX}C_YC_X.$$

4.1. Bias and MSE of $\overline{y}_{ESL1}$

Let $k = \frac{1}{\ln \overline{X}}$. Now we express the logarithm of the sample mean of size r, $\overline{x}_r$ as

$$\ln \overline{x}_r = \ln \overline{X}(1 + e_2) = \ln \overline{X} + \ln(1 + e_2).$$

Assuming that $|e_2| < 1$, we expand $\ln(1 + e_2)$ up to second-order approximation as

$$\ln \overline{x}_r = \ln \overline{X} + \left(e_2 - \frac{e_2^2}{2} + \frac{e_2^3}{3} - \frac{e_2^4}{4} + \ldots\right) \approx \ln \overline{X}\left(1 + ke_2 - \frac{ke_2^2}{2}\right). \tag{4}$$

Using Equation (4) in Equation (1), expanding the proposed estimator $\overline{y}_{\text{ESL1}}$ up to second order in terms of e_i's, we have

$$\overline{y}_{\text{ESL1}} \approx \overline{Y}\left[1 + e_1 + \alpha\beta k e_2 - \frac{\alpha\beta k(1 + 2k + \beta k - k - \alpha\beta k)}{2} e_2^2 + \alpha\beta k e_1 e_2\right]. \tag{5}$$

Taking expectations on both sides of Equation (5) and rewritten as

$$E\left(\overline{y}_{\text{ESL1}} - \overline{Y}\right) = \overline{Y} E\left[\alpha\beta k e_1 e_2 - \frac{\alpha\beta k(1 + 2k + \beta k - k - \alpha\beta k)}{2} e_2^2\right]. \tag{6}$$

Now, using the results of $E\left(e_2^2\right)$ and $E\left(e_1 e_2\right)$ in Equation (6), the bias of $\overline{y}_{\text{ESL1}}$ is

$$B\left(\overline{y}_{\text{ESL1}}\right) = \alpha\beta k \overline{Y} \lambda_{r,N}\left(\phi_{YX} - \frac{1 + 2k + \beta k - k - \alpha\beta k}{2}\right) C_X^2. \tag{7}$$

Squaring both sides of the Equation (5) and taking expectations up to second order, we have

$$E\left(\overline{y}_{\text{ESL1}} - \overline{Y}\right)^2 = \overline{Y}^2 E\left[e_1^2 + \alpha^2\beta^2 k^2 e_2^2 + 2\alpha\beta k e_1 e_2\right]. \tag{8}$$

Putting the results of $E\left(e_1^2\right)$, $E\left(e_2^2\right)$ and $E\left(e_1 e_2\right)$ in Equation (8), the MSE of $\overline{y}_{\text{ESL1}}$ is given by

$$\text{MSE}\left(\overline{y}_{\text{ESL1}}\right) = \overline{Y}^2 \lambda_{r,N}\left[C_Y^2 + \alpha\beta k(\alpha\beta k + 2\phi_{YX}) C_X^2\right]. \tag{9}$$

Differentiating Equation (9) with respect to α and equating the derivative to zero, the optimal value of α is

$$\alpha_{\text{opt}} = -\frac{\phi_{YX}}{\beta k}.$$

Putting the optimal value of α in Equation (9), we get the asymptotic optimum mean square error of the estimator $\left(\overline{y}_{\text{ESL1}}\right)$ as

$$\text{MSE}\left(\overline{y}_{\text{ESL1}}\right)_{\text{opt}} = \overline{Y}^2 \lambda_{r,N}\left(1 - \rho_{YX}^2\right) C_Y^2. \tag{10}$$

4.2. Bias and MSE of $\overline{y}_{\mathbf{ESL2}}$

Let $k = \frac{1}{\ln \overline{X}}$. Now we express the logarithm of the sample mean of size n, $\overline{x}_n$ as

$$\ln \overline{x}_n = \ln \overline{X}(1 + e_3) = \ln \overline{X} + \ln(1 + e_3).$$

Assuming that $|e_3| < 1$, we expand $\ln(1 + e_3)$ up to second-order approximation as

$$\ln \overline{x}_n = \ln \overline{X} + \left(e_3 - \frac{e_3^2}{2} + \frac{e_3^3}{3} - \frac{e_3^4}{4} + \ldots\right) \approx \ln(\overline{X})\left(1 + ke_3 - \frac{ke_3^2}{2}\right). \tag{11}$$

Using Equation (11) in Equation (2), expanding the proposed estimator $\overline{y}_{\text{ESL2}}$ up to second order in terms of e_i's, we have

$$\overline{y}_{\text{ESL2}} \approx \overline{Y}\left[1 + e_1 + \alpha\beta k e_3 - \frac{\alpha\beta k(1 + 2k + \beta k - k - \alpha\beta k)}{2} e_3^2 + \alpha\beta k e_1 e_3\right]. \tag{12}$$

Taking expectations on both sides of Equation (12), we rewrite it as

$$E\left(\overline{y}_{\text{ESL2}} - \overline{Y}\right) = \overline{Y} E\left[\alpha\beta k e_1 e_3 - \frac{\alpha\beta k(1 + 2k + \beta k - k - \alpha\beta k)}{2} e_3^2\right]. \tag{13}$$

Now, using the results of $E\left(e_3^2\right)$ and $E\left(e_1 e_3\right)$ in Equation (13), the bias of $\overline{y}_{\text{ESL2}}$ is

$$B\left(\overline{y}_{\text{ESL2}}\right)=\alpha\beta k\overline{Y}\lambda_{n,N}\left(\phi_{YX}-\frac{1+2k+\beta k-k-\alpha\beta k}{2}\right)C_X^2. \tag{14}$$

Squaring both sides of Equation (12) and taking expectations up to second order, we have

$$E\left(\overline{y}_{\text{ESL2}}-\overline{Y}\right)^2=\overline{Y}^2E\left[e_1^2+\alpha^2\beta^2k^2e_3^2+2\alpha\beta ke_1e_3\right]. \tag{15}$$

Substituting the expressions for $E\left(e_1^2\right)$, $E\left(e_3^2\right)$ and $E\left(e_1e_3\right)$ in Equation (15), the MSE of $\overline{y}_{\text{ESL2}}$ is given by

$$\text{MSE}\left(\overline{y}_{\text{ESL2}}\right)=\overline{Y}^2\left[\lambda_{r,N}C_Y^2+\alpha\beta k\lambda_{n,N}\left(\alpha\beta k+2\phi_{YX}\right)C_X^2\right]. \tag{16}$$

Differentiating Equation (16) with respect to α and equating the derivative to zero, the optimal value of α is

$$\alpha_{\text{opt}}=-\frac{\phi_{YX}}{\beta k}.$$

Substituting the optimal value of α in Equation (16), we get the asymptotic optimum mean square error of the suggested estimator $(\overline{y}_{\text{ESL2}})$ as

$$\text{MSE}\left(\overline{y}_{\text{ESL2}}\right)_{\text{opt}}=\overline{Y}^2\left(\lambda_{r,N}-\lambda_{n,N}\rho_{YX}^2\right)C_Y^2. \tag{17}$$

4.3. Bias and MSE of $\overline{y}_{\text{ESL3}}$

Using Equations (4) and (11) in Equation (3) and expanding the proposed estimator $\overline{y}_{\text{ESL3}}$ up to second order in terms of e_i's, we have

$$\begin{aligned}\overline{y}_{\text{JLE3}}\approx\overline{Y}\Big[&1+e_1+\alpha\beta ke_2-\alpha\beta ke_3+\frac{\alpha\beta k}{2}\left(1-\beta k+k+\alpha\beta k\right)e_3^2\\&-\frac{\alpha\beta k}{2}\left(1+k+\beta k-\alpha\beta k\right)e_2^2+\alpha\beta ke_1e_2-\alpha\beta ke_1e_3+\alpha\beta^2k^2(1-\alpha)e_2e_3\Big].\end{aligned} \tag{18}$$

Taking expectations on both sides of (18) and rewritten as

$$\begin{aligned}E\left(\overline{y}_{\text{ESL3}}-\overline{Y}\right)=\overline{Y}E\Big[&\frac{\alpha\beta k}{2}\left(1-\beta k+k+\alpha\beta k\right)e_3^2-\frac{\alpha\beta k}{2}\left(1+k+\beta k-\alpha\beta k\right)e_2^2\\&+\alpha\beta ke_1e_2-\alpha\beta ke_1e_3+\alpha\beta^2k^2(1-\alpha)e_2e_3\Big].\end{aligned} \tag{19}$$

Now using the results of $E\left(e_1^2\right)$, $E\left(e_2^2\right)$, $E\left(e_3^2\right)$, $E\left(e_2e_3\right)$, $E\left(e_1e_3\right)$ and $E\left(e_1e_2\right)$ in (19), the bias of $\overline{y}_{\text{ESL3}}$ is

$$B\left(\overline{y}_{\text{ESL3}}\right)=\overline{Y}\frac{\alpha\beta k}{2}\lambda_{r,n}\left[2\phi_{YX}-\left(1+\beta k+k-\alpha\beta k\right)\right]C_X^2. \tag{20}$$

Squaring both sides of Equation (18) and taking expectations up to second order, we have

$$E\left(\overline{y}_{\text{ESL3}}-\overline{Y}\right)^2=\overline{Y}^2E\left[e_1^2+\alpha^2\beta^2k^2e_2^2+\alpha^2\beta^2k^2e_3^2-2\alpha^2\beta^2k^2e_2e_3-2\alpha\beta ke_1e_3+2\alpha\beta ke_1e_2\right]. \tag{21}$$

Substituting the results of $E\left(e_1^2\right)$, $E\left(e_2^2\right)$, $E\left(e_3^2\right)$, $E\left(e_2e_3\right)$, $E\left(e_1e_3\right)$ and $E\left(e_1e_2\right)$ in (21), the MSE of $\overline{y}_{\text{ESL3}}$ is obtained as

$$\text{MSE}\left(\overline{y}_{\text{ESL3}}\right)=\overline{Y}^2\left[\lambda_{r,N}C_Y^2+\lambda_{r,n}\alpha\beta k\left(\alpha\beta k+2\phi_{YX}\right)C_X^2\right]. \tag{22}$$

Differentiating Equation (22) with respect to α and equating the derivative to zero, the optimal value of α is

$$\alpha_{\text{opt}}=-\frac{\phi_{YX}}{\beta k}.$$

Substituting the expressions for α in (22), we get the asymptotic optimum mean square error of the suggested estimator $(\overline{y}_{\text{ESL3}})$ as

$$\text{MSE}\left(\overline{y}_{\text{ESL3}}\right)_{\text{opt}}=\overline{Y}^2\left(\lambda_{r,N}-\lambda_{r,n}\rho_{YX}^2\right)C_Y^2. \tag{23}$$

5. Efficiency comparison of the estimator $\overline{y}_{\mathrm{ESL1}}$

Versus the estimator $\overline{y}_r$,

$$\mathrm{MSE}\left(\overline{y}_r\right)-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL1}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{n,N}\rho_{YX}^2C_Y^2>0.$$

Hence the proposed class of estimator is more competent than $\overline{y}_r$.

Versus the estimator $\overline{y}_{\mathrm{RAT}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{RAT}}\right)-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL1}}\right)_{\mathrm{opt}}=\overline{Y}^2\left[\lambda_{n,N}\rho_{YX}^2C_Y^2+\lambda_{r,n}\left(C_X-\rho_{YX}C_Y\right)^2\right]>0.$$

Hence the proposed class of estimator is more competent than $\overline{y}_{\mathrm{RAT}}$.

Versus the estimator $\overline{y}_{\mathrm{COMP}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{COMP}}\right)_{\mathrm{opt}}-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL1}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{n,N}\rho_{YX}^2C_Y^2>0.$$

Hence the proposed estimator $\overline{y}_{\mathrm{ESL1}}$ is more competent than the existing estimator $\overline{y}_{\mathrm{COMP}}$.

6. Efficiency comparison of the estimator $\overline{y}_{\mathrm{ESL2}}$

Versus the estimator $\overline{y}_r$,

$$\mathrm{MSE}\left(\overline{y}_r\right)-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL2}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{n,N}\rho_{YX}^2C_Y^2>0.$$

Hence the proposed class of estimator is more competent than $\overline{y}_r$.

Versus the estimator $\overline{y}_{\mathrm{RAT}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{RAT}}\right)-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL2}}\right)_{\mathrm{opt}}=\overline{Y}^2\left[\left(\lambda_{n,N}-\lambda_{r,n}\right)\rho_{YX}^2C_Y^2+\lambda_{r,n}\left(C_X-\rho_{YX}C_Y\right)^2\right]>0.$$

Hence the proposed estimator $\overline{y}_{\mathrm{ESL2}}$ is more competent than the estimator $\overline{y}_{\mathrm{RAT}}$ under the condition of $r>\frac{Nn}{2N-n}$.

Versus the estimator $\overline{y}_{\mathrm{COMP}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{COMP}}\right)_{\mathrm{opt}}-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL2}}\right)_{\mathrm{opt}}=\overline{Y}^2\left(\lambda_{n,N}-\lambda_{r,n}\right)\rho_{YX}^2C_Y^2>0.$$

Hence the estimator $\overline{y}_{\mathrm{ESL2}}$ is better than the estimator $\overline{y}_{\mathrm{COMP}}$ if $r>\frac{Nn}{2N-n}$.

Versus the estimator $\overline{y}_{\mathrm{ESL1}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL2}}\right)_{\mathrm{opt}}-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL1}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{r,n}\rho_{YX}^2C_Y^2>0.$$

Hence the estimator $\overline{y}_{\mathrm{ESL1}}$ is more efficient than $\overline{y}_{\mathrm{ESL2}}$.

7. Efficiency comparison of the estimator $\overline{y}_{\mathrm{ESL3}}$

Versus the estimator $\overline{y}_r$,

$$\mathrm{MSE}\left(\overline{y}_r\right)-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL3}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{r,n}\rho_{YX}^2C_Y^2>0.$$

It is found that the estimator is more competent than $\overline{y}_r$.

Versus the estimator $\overline{y}_{\mathrm{RAT}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{RAT}}\right)-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL3}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{r,n}\left(C_X-\rho_{YX}C_Y\right)^2>0.$$

Hence the proposed estimator $\overline{y}_{\mathrm{ESL3}}$ is more competent than the existing estimator $\overline{y}_{\mathrm{RAT}}$.

Versus the estimator $\overline{y}_{\mathrm{COMP}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{COMP}}\right)_{\mathrm{opt}}-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL3}}\right)_{\mathrm{opt}}=0.$$

Hence the estimators $\overline{y}_{\mathrm{ESL3}}$ and $\overline{y}_{\mathrm{COMP}}$ are equally efficient.

Versus the estimator $\overline{y}_{\mathrm{ESL1}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL3}}\right)_{\mathrm{opt}}-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL1}}\right)_{\mathrm{opt}}=\overline{Y}^2\lambda_{n,N}\rho_{YX}^2C_Y^2>0.$$

Hence the estimator $\overline{y}_{\mathrm{ESL1}}$ is more efficient than $\overline{y}_{\mathrm{ESL3}}$.

Versus the estimator $\overline{y}_{\mathrm{ESL2}}$,

$$\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL3}}\right)_{\mathrm{opt}}-\mathrm{MSE}\left(\overline{y}_{\mathrm{ESL2}}\right)_{\mathrm{opt}}=\overline{Y}^2\left(\lambda_{n,N}-\lambda_{r,n}\right)\rho_{YX}^2C_Y^2>0.$$

Hence the estimator $\overline{y}_{\mathrm{ESL2}}$ is more efficient than $\overline{y}_{\mathrm{ESL3}}$ if $r>\frac{Nn}{2N-n}$.

8. Numerical illustration

To evaluate the performance and efficiency of the proposed estimators, we selected two natural population data sets. The details of the data sets are given below

Population A. Source: District Census Handbook, Orissa (1981). We considered population data of 109 villages/towns/wards in the urban area under police-station Baria, Tahasil-Champua, Orissa, where

$$X = \text{Average number of non-workers in the village},$$
$$Y = \text{Average number of literate persons in the village}.$$

Population B. We considered COVID-19 death data in India. COVID-19 data were retrieved from WHO websites (download link: https://covid19.who.int/WHO-COVID-19-global-data.csv) (2022). A total of 943 days' data (from the period of 01-February-2020 to 31-August-2022) were taken to examine the impact of mortality in India.

Table 1. Descriptions of the population data sets.

Parameters	Population A	Population B
N	109	943
n (10% − 45%)	(16, 27, 31)	(330, 280, 308)
r (75% − 95%)	(8, 12, 14), (13, 19, 24), (19, 22, 28)	(264, 272, 298), (230, 246, 268), (222, 256, 272)
$\overline{Y}$	145.3028	559.7815
$\overline{X}$	259.083	47113.88
ρ	0.875	0.7393
C_Y^2	0.4759	2.3910
C_X^2	0.2947	2.6354

Here, we have considered the sample sizes (n) between 10% and 45% and the response rate (r) between 75% and 95% with different correlation coefficients.

The percentage relative efficiencies (PREs) of the proposed estimator and exiting estimators with respect to the mean imputation estimator ($\overline{y}_r$) have been calculated for different choices of n and r and presented in Tables 2 and 3. The expressions of PRE are given by

$$\text{PRE}(t) = 100 \times \frac{\text{MSE}\,(\overline{y}_r)}{\text{MSE}(t)},$$

where $t = \overline{y}_{\text{RAT}}, \overline{y}_{\text{COMP}}, \overline{y}_{\text{ESL1}}, \overline{y}_{\text{ESL2}}, \overline{y}_{\text{ESL3}}$.

We have computed PRE of different estimators under population data sets.

Table 2. PRE of the proposed and existing estimators with respect to mean for population data set A.

n	r	Mean ($\overline{y}_r$)	Ratio ($\overline{y}_{RAT}$)	Compromised ($\overline{y}_{COMP}$)	$\overline{y}_{\text{ESL1}}$	$\overline{y}_{\text{ESL2}}$	$\overline{y}_{\text{ESL3}}$
16	8	100	133.4421	170.3967	426.6667	194.4379	170.3967
	12	100	115.0050	127.4024	426.6667	222.4890	127.4024
	14	100	107.1363	112.3352	426.6667	290.5442	112.3352
27	13	100	137.6329	182.0664	426.6667	245.9588	182.0664
	19	100	119.9993	137.8821	426.6667	196.4182	137.8821
	24	100	107.0864	112.2447	426.6667	291.1513	112.2447
31	19	100	127.8340	155.9912	426.6667	131.8468	155.9912
	22	100	120.3272	138.5975	426.6667	194.9844	138.5975
	28	100	106.4376	111.0747	426.6667	299.3295	111.0747

Table 3. PRE of the proposed and existing estimators with respect to mean for population data set B.

n	r	Mean ($\overline{y}_r$)	Ratio ($\overline{y}_{\text{RAT}}$)	Compromised ($\overline{y}_{\text{COMP}}$)	$\overline{y}_{\text{ESL1}}$	$\overline{y}_{\text{ESL2}}$	$\overline{y}_{\text{ESL3}}$
330	264	100	114.29	117.8987	220.5385	165.221	117.8987
	272	100	112.5097	115.6074	220.5387	169.9412	112.5097
	298	100	106.8168	108.3996	220.5387	188.3514	106.8168
280	230	100	111.8960	114.8218	220.5387	171.6678	111.8960
	246	100	107.9858	109.8651	220.5385	184.0847	109.8651
	268	100	102.7698	103.3832	220.5387	205.6936	102.7698
308	222	100	119.6731	124.9379	220.5387	153.1305	119.6731
	256	100	111.6468	114.5032	220.5385	172.3846	114.5033
	272	100	107.9846	109.8636	220.5387	184.0888	107.9846

9. Simulation study

In this section, we conduct a simulation study using statistical computational software [9] to establish the performance of the proposed imputation methods over mean and ratio methods of imputations. For this manuscript, we have generated data sets using the 'genCorGen' function available in the package 'simstudy' [10]. For the study and auxiliary variables, we generated data sets from the Normal distribution with given parameters and correlation coefficients.

We consider a population size of $N = 8500$ to generate a dataset of the variables (X, Y) and calculate the required parameters and correlation coefficient, and we consider only one combination for simulation for $\beta = 2$.

For data from the normal distribution:

$$Data = (X, Y), \quad X \sim (321.0029, 0.9986), \quad Y \sim (457.0018, 0.9925).$$

The following steps were used for the simulation of the required population:

Step 1 The 'genCorGen' function was used to create a data set with the normal distribution of variables X and Y of size $N = 8500$ using the statistical computer program [9] Software.

Step 2 The parameters were calculated.

Step 3 A sample of sizes $n = 340, 765, 1020, 1445, 1700$, and 1870 (i.e., sample sizes lay between 4% and 22%) was chosen from this artificial data set with response rates $r = (265, 282, 296, 313)$, $(597, 635, 666, 704)$, $(796, 847, 887, 939)$, $(1127, 1199, 1257, 1329)$, $(1326, 1411, 1479, 1564)$, and $(1459, 1552, 1627, 1720)$, respectively (i.e., the response rate lay between 78% and 92% of the total).

Step 4 Sample statistics, i.e. sample mean, sample variance, and the values of the proposed and existing estimators were calculated for this sample by the imputation techniques.

Step 5 Steps (3) and (4) were repeated 50 000 times.

Step 6 The MSE of every estimator was calculated by the formula of the MSE given by

$$\text{MSE}(t) = \frac{1}{50000} \sum_{i=1}^{m} \left(t_i - \overline{Y}\right)^2,$$

where $t = \overline{y}_{\text{RAT}}, \overline{y}_{\text{COMP}}, \overline{y}_{\text{ESL1}}, \overline{y}_{\text{ESL2}}, \overline{y}_{\text{ESL3}}$.

Step 7 The PRE of each estimator was calculated with respect to both mean and ratio estimators.

Table 4. PRE of the proposed and existing estimators w.r.t. mean for simulated population.

n	r	Mean ($\overline{y}_r$)	Ratio ($\overline{y}_{\text{RAT}}$)	Compromised ($\overline{y}_{\text{COMP}}$)	$\overline{y}_{\text{ESL1}}$	$\overline{y}_{\text{ESL2}}$	$\overline{y}_{\text{ESL3}}$
340	265	100	113.9510	117.9843	298.4642	205.8361	117.9844
	282	100	110.3634	113.1978	296.6979	220.1958	113.1977
	296	100	107.6579	109.5184	298.1828	235.3574	109.5187
	313	100	104.5903	105.7345	302.5664	260.0924	105.7344
765	597	100	114.2429	118.0520	299.3358	202.9238	118.0520
	635	100	111.7084	114.6441	300.7902	221.9763	114.6441
	666	100	108.3124	110.4426	296.4110	235.3096	110.4426
	704	100	104.8390	106.1487	296.3235	251.2221	106.1487
1020	794	100	114.3723	118.8043	303.0546	203.9413	118.8043
	847	100	111.6880	114.3572	299.4302	218.4069	114.3572
	887	100	108.5921	110.5954	303.0168	234.8558	110.5954
	939	100	105.5798	106.8347	295.7327	251.2449	106.8347
1445	1127	100	117.3812	122.0280	301.8481	200.6764	122.0280
	1199	100	111.7990	114.9237	297.0182	211.8166	114.9237
	1257	100	108.8625	111.3301	300.3498	228.1987	111.3301
	1329	100	105.1634	106.6227	295.3727	248.1224	106.6227
1700	1326	100	116.8813	121.6795	299.3749	196.6359	121.6795
	1411	100	111.4307	115.0038	299.1192	212.0901	115.0038
	1479	100	109.8892	112.4304	293.5836	224.4231	112.4304
	1564	100	105.7665	107.3481	293.5836	224.4231	112.4304
1870	1459	100	115.7670	120.4316	294.2983	192.2403	120.4316
	1552	100	113.3495	116.2517	300.6417	212.6468	116.2517
	1627	100	109.7758	112.1398	294.2248	224.0924	112.1398
	1720	100	105.7008	107.2563	299.8213	250.3383	107.2563

Table 5. PRE of the proposed and existing estimators w.r.t. ratio for simulated population.

n	r	Mean ($\overline{y}_r$)	Ratio ($\overline{y}_{\text{RAT}}$)	Compromised ($\overline{y}_{\text{COMP}}$)	$\overline{y}_{\text{ESL1}}$	$\overline{y}_{\text{ESL2}}$	$\overline{y}_{\text{ESL3}}$
340	265	87.7570	100	103.5395	261.9233	180.6356	103.5396
	282	90.6097	100	102.5682	268.8372	199.5188	102.5682
	296	92.8868	100	101.7284	276.9726	218.6161	101.7284
	313	95.6112	100	101.0940	289.2873	248.6774	101.0939
765	597	87.5328	100	103.3342	262.0169	177.6248	103.3342
	635	89.5188	100	102.6279	269.2637	198.7104	102.6279
	666	92.3256	100	101.9668	273.6631	217.2509	101.9668
	704	95.3844	100	101.2492	282.6464	239.6267	101.2492
1020	794	87.4337	100	103.8751	264.9719	178.3135	103.8751
	847	89.5351	100	102.3898	268.0953	195.5509	102.3898
	887	92.0878	100	101.8448	279.0414	216.2734	101.8448
	939	94.7151	100	101.1886	280.1035	237.9669	101.1886
1445	1127	85.1925	100	103.9587	257.1519	170.9612	103.9587
	1199	89.4463	100	102.7949	265.6717	189.4620	102.7949
	1257	91.8589	100	102.2667	175.8983	209.6210	102.2667
	1329	95.0901	100	101.3876	280.8703	235.9398	101.3876
1700	1326	85.5569	100	104.1052	256.1359	168.2356	104.1052
	1411	89.7419	100	103.2066	268.4353	190.3337	103.2066
	1479	91.0008	100	102.3125	267.1632	204.2267	102.3125
	1564	94.5479	100	101.4954	277.2213	233.6472	101.4954
1870	1459	86.3804	100	104.0293	254.2159	166.0578	104.0293
	1552	88.2227	100	102.5604	265.2343	187.6028	102.5604
	1627	91.0948	100	102.1535	268.0235	204.1364	102.1535
	1720	94.6067	100	101.4716	283.6509	236.8367	101.4716

10. Conclusion

In this work, we have considered both real and simulated data to conclude:

— Based on three population data sets considered to examine, we observe that the proposed estimators $\overline{y}_{\text{ESL1}}$ and $\overline{y}_{\text{ESL2}}$ are more efficient than the existing estimators such as mean, ratio, and compromised; and the third proposed estimator $\overline{y}_{\text{ESL3}}$ is better than mean and ratio; and equally efficient as compromised estimator.

— We generated simulated data sets using statistical software R from the normal distribution and conducted a simulation to examine the efficiencies of the proposed estimators under mean and ratio methods for different values of sample size and response rate. We observe from Tables 4–5 that the results of the efficiencies of the proposed estimators have the same trend as that of Tables 2–3.

When all the sample unit data are accessible, estimators perform better. In the case of a lack of availability, due to non-response in single-phase sampling, the imputation technique becomes effective. Our proposed estimators are limited to handling non-response in single-phase studies only.

[1] Meeden G. A decision theoretic approach to imputation in finite population sampling. Journal of the American Statistical Association. **95** (450), 586–595 (2000).

[2] Singh S., Horn S. Compromised imputation in survey sampling. Metrika. **51** (3), 267–276 (2000).

[3] Singh A. K., Singh P., Singh V. K. Exponential-Type Compromised Imputation in Survey Sampling. Journal of Statistics Applications & Probability. **3** (2), 211–217 (2014).

[4] Prasad S. Some product exponential methods of imputation in sample surveys. Statistics in Transition new series. **19** (1), 159–166 (2018).

[5] Singh G. N., Bhattacharyya D., Bandyopadhyay A. Some logarithmic and sine-type imputation techniques for missing data in survey sampling in the presence of measurement errors. Journal of Statistical Computation and Simulation. **91** (4), 713–731 (2020).

[6] Pandey R., Thakur N. S., Yadav K. Estimation of population mean using exponential ratio type imputation method under survey non-response. Journal of the Indian Statistical Association. **53** (1&2), 89–107 (2015).

[7] Singh B. K., Gogoi U. Estimation of Population mean using Exponential Dual to Ratio Type Compromised Imputation for Missing data in Survey Sampling. Journal of Statistics Applications & Probability. **6** (3), 515–522 (2017).

[8] Lee H., Rancourt E., Särndal C. E. Experiments with Variance Estimation from Survey Data with Imputed Values. Journal of Official Statistics. **10** (3), 231–243 (1994).

[9] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna (2022). `https://www.R-project.org`.

[10] Goldfeld K., Wujciak-Jens J. Simstudy: Simulation of Study Data. `https://cran.r-project.org/web/packages/simstudy/index.html`.

Вивчення деяких методів експоненціально-оберненої синусо-логарифмічної імпутації за умов відсутніх даних з використанням змодельованих даних

Сінгх Б. К., Дас Дж.

Математичний факультет,
Північно-Східний регіональний інститут науки і технологій,
Нірджулі, Аруначал-Прадеш 791109, Індія

Відсутність відповіді в опитуванні виникає через помітні відмінності між людьми, які вирішили взяти участь у певному опитуванні, та тими, хто цього не зробив, що може призвести до неповноти даних. Метод імпутації є одним із найкращих способів боротьби з відсутністю відповідей у вибірковому опитуванні для забезпечення повноти даних. Найпоширенішими методами імпутації є імпутація середнього значення, імпутація за співвідношенням та метод компромісної імпутації. Імпутація за середнім значенням замінює всі пропущені значення середнім значенням відповідей, що зменшує варіативність у наборі даних. Метод імпутації за співвідношенням є досить корисним для підтримки пропорційного зв'язку між змінними. Він передбачає сильний лінійний зв'язок між досліджуваною та допоміжною змінними, що не завжди виконується, і в разі порушення цього припущення метод імпутації може бути зміщеним. Метод компромісної імпутації, що поєднує різні методи, має кілька обмежень і все ще може давати зміщені результати, якщо основні припущення окремих методів порушуються. Через обмеження традиційних методів імпутації та для їх подолання було запропоновано три експоненційно-обернені синусно-логарифмічні методи імпутації та відповідні точкові оцінки. Введено зміщення (bias) та середньоквадратичну помилку (MSE) запропонованих оцінок і встановлено теоретичні та числові результати їхньої ефективності порівняно з існуючими оцінками. Також згенеровано імітаційні набори даних за допомогою статистичного програмного забезпечення для проведення симуляційного дослідження. Крім того, для дослідження ефективності запропонованих оцінок було розраховано відсоткову відносну ефективність (PRE) запропонованих та існуючих оцінок, використовуючи імітаційні набори даних щодо методів імпутації за середнім значенням та за співвідношенням. Аналіз результатів показує, що запропоновані методи імпутації є більш ефективними, ніж існуючі.

Ключові слова: *простий випадковий відбір; імпутація; зміщення; середньоквадратична похибка (MSE); відсоткова відносна ефективність (PRE); симуляційне дослідження.*