Generalized Class of Estimators for Mean of Population through Imputation Technique using SRSWOR

*

Abstract

In the current study, we have developed two generalized classes of estimators for estimating the mean of population based on imputation technique under simple random sampling without replacement (SRSWOR), and their statistical properties have been derived. Traditional imputation techniques such as mean and ratio methods have been found to be the particular cases of the developed imputation techniques. Also, some estimators discussed in the literature have been found as the members of the developed generalized class of estimators. Developed estimators outperform under the given mathematical conditions. It is then verified with the help of empirical study using two authentic datasets.

Keywords: Estimators; Imputation; Bias; Missing Data

1 Introduction

Missing completely at random (MCAR) occurs when the absence of data is completely unrelated to both the observed data (the data that is present) and the unobserved data (data that is missing). Occasionally we obtain data with missing completely at random in sample surveys. Using such data to estimate the mean of the population does not give an appropriate estimate. In those circumstances, imputation techniques have been used to estimate the mean for the population. These techniques have been used to enhance the precision and truthfulness of data analysis. Many imputation techniques, including the mean, ratio, regression and power transformation methods have been utilized to substitute missing observations.

The imputation technique based on ancillary variable was initiated by Lee et al.(1994). Further, Singh and Horn(2000), Singh and Deo(2003), Ahmed et al.(2006), Prasad(2018), Bhushan and Pandey(2018), Bhushan et al.(2023) including others have formed a variety of imputation techniques based on information on ancillary variable. Recent advancements for population mean estimation in the case of missing data have been done by Pandey et al.(2025), Sinsomboonthong and Sinsomboonthong (2025).

In the current research, we have developed two generalized classes of estimators for estimating the mean of population based on imputation technique under simple random sampling without replacement (SRSWOR), by assuming the missing completely at random (MCAR) technique to impute the missing data and their statistical properties have been derived. Some of the estimators existing in literature have been found as the particular cases of the developed estimators. Developed generalized classes of estimators outperform under the given mathematical conditions. Empirical analysis using two authentic data sets have been done to prove the dominance of the developed estimators over other corresponding estimators found as their special cases.

2 Terminologies and sampling method used

Let the population Υ=$\{1, 2, ..., N\}$ of size N is considered to be finite. Again, let ($\overline{Y}$, $\overline{X}$) denote the population mean for N observations on study variable (y) and ancillary variable (x). For the estimation of population mean $\overline{Y}$, a random sample s of size n is chosen from Υ by using simple random sampling without replacement (SRSWOR). Further, let r denotes the number of responding units taken from n units and A indicates the set of units that respond. Also, A^c represents the set of non-responding units. For units $j \in A$, values of sample on study variables y have been observed while for missing values as $j \in A^c$, suitable values are acquired. It has been supposed that through the help of an ancillary variable x, the imputation technique is carried out and $x_s = \{x_j : j \in s\}$ are known.

Let $\overline{y}_r = \frac{1}{r}\sum\limits_{j=1}^{r} y_j$, $\overline{y}_n = \frac{1}{n}\sum\limits_{j=1}^{n} y_j$ denote the response mean and sample mean for study variable y and $\overline{x}_r = \frac{1}{r}\sum\limits_{j=1}^{r} x_j$, $\overline{x}_n = \dfrac{1}{n}\sum\limits_{j=1}^{n} x_j$ denote the response mean and sample mean for ancillary variable x.

The expressions of population mean square, population covariance and coefficient of variation for Y and X are given as follows:
$S_y^2 = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(y_j - \overline{Y}\right)^2$: Population mean square of study variable y,
$S_x^2 = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(x_j - \overline{X}\right)^2$: Population mean square of ancillary variable x,

$S_{yx} = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(y_j - \overline{Y}\right)\left(x_j - \overline{X}\right)$: Population covariance between study variable y and ancillary variable x,
$C_y = \frac{S_y}{\overline{Y}}$: Coefficient of variation of study variable y,
$C_x = \frac{S_x}{\overline{X}}$: Coefficient of variation of ancillary variable x.

In order to acquire the mean square error (MSE) of the developed generalized class of imputation methods, let $\overline{y}_r = \overline{Y}(1+\epsilon_0)$, $\overline{x}_r = \overline{X}(1+\epsilon_1)$, $\overline{x}_n = \overline{X}(1+\epsilon_2)$ such that $E(\epsilon_0) = 0, E(\epsilon_1) = 0, E(\epsilon_2) = 0$, and $E(\epsilon_0^2) = f_r C_y^2$, $E(\epsilon_1^2) = f_r C_x^2$, $E(\epsilon_2^2) = f_n C_x^2$, $E(\epsilon_0\epsilon_1) = f_r \rho C_y C_x$, $E(\epsilon_1\epsilon_2) = f_n C_x^2$ and $E(\epsilon_0\epsilon_2) = f_n \rho C_y C_x$, where ρ is the coefficient of correlation between study and ancillary variables, $f_r = \left(\frac{1}{r} - \frac{1}{N}\right)$ and $f_n = \left(\frac{1}{n} - \frac{1}{N}\right)$.

3 Suggested Imputation Methods

The objective of the current paper is to extend the research work of Ahmed et al. (2006) by developing the estimators for population mean in case of missing completely at random. Keeping this in view, we have developed the following generalized class of estimator for the mean of population using ($\overline{X}$, $\overline{x}_n$) and imputation method as:

$$y_{.j1} = \begin{cases} y_j & \text{if } j \in A \\ \dfrac{1}{n-r}\left[n\overline{y}_r\left(\dfrac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha} - r\overline{y}_r\right] & \text{if } j \in A^c \end{cases} \tag{3.1}$$

Estimator for the mean of population using above imputation method has been derived as:

$$\begin{aligned} {P_1}^* &= \frac{1}{n}\sum_{j=1}^{n} y_{.j1} = \frac{1}{n}\left[\sum_{j\in A} y_{.j1} + \sum_{j\in A^c} y_{.j1}\right] \\ &= \frac{1}{n}\left[\sum_{j=1}^{r} y_j + \sum_{j=1}^{n-r}\frac{1}{n-r}\left\{n\overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha} - r\overline{y}_r\right\}\right] \\ &= \frac{1}{n}\left[r\overline{y}_r + (n-r)\frac{1}{n-r}\left\{n\overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha} - r\overline{y}_r\right\}\right] \\ &= \frac{1}{n}\left[r\overline{y}_r + n\overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha} - r\overline{y}_r\right] \\ &= \frac{1}{n}\left[n\overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha}\right] = \overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha} \end{aligned} \tag{3.2}$$

Hence, estimator for the mean of population ($\overline{Y}$) has been obtained as:

$${P_1}^* = \overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha}, \tag{3.3}$$

where $0 \leq q_1 \leq 1$ and α is a chosen constant.

Now by using information on $\overline{X}$ and $\overline{x}_r$, we have developed another generalized class of estimator for the mean of population using imputation method as:

$$y_{.j2} = \begin{cases} y_j & \text{if } j \in A \\ \dfrac{1}{n-r}\left[n\overline{y}_r\left(\dfrac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta} - r\overline{y}_r\right] & \text{if } j \in A^c \end{cases} \tag{3.4}$$

By using the above imputation technique, estimator for the mean of population has been derived as:

$$
\begin{aligned}
{P_2}^* &= \frac{1}{n}\sum_{j=1}^{n} y_{.j2} = \frac{1}{n}\left[\sum_{j\in A} y_{.j2} + \sum_{j\in A^c} y_{.j2}\right] \\
&= \frac{1}{n}\left[\sum_{j=1}^{r} y_j + \sum_{j=1}^{n-r}\frac{1}{n-r}\left\{n\overline{y}_r\left(\frac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta} - r\overline{y}_r\right\}\right] \\
&= \frac{1}{n}\left[r\overline{y}_r + (n-r)\frac{1}{n-r}\left\{n\overline{y}_r\left(\frac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta} - r\overline{y}_r\right\}\right] \\
&= \frac{1}{n}\left[r\overline{y}_r + n\overline{y}_r\left(\frac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta} - r\overline{y}_r\right] \\
&= \frac{1}{n}\left[n\overline{y}_r\left(\frac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta}\right] = \overline{y}_r\left(\frac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta} \qquad (3.5)
\end{aligned}
$$

Hence, estimator for the mean of population ($\overline{Y}$) has been obtained as:

$$
{P_2}^* = \overline{y}_r\left(\frac{q_2\overline{x}_r + (1-q_2)\overline{X}}{\overline{X}}\right)^{\beta}, \qquad (3.6)
$$

where $0 \leq q_2 \leq 1$ and β is a chosen constant.

4 Developed Estimators' Bias and Mean Square Error(MSE)

Considering the developed generalized class of estimator ${P_1}^* = \overline{y}_r\left(\frac{q_1\overline{x}_n + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha}$.

Making use of the notations specified in section **2**, the estimator previously indicated has been expressed as:

$$
\begin{aligned}
{P_1}^* &= \overline{Y}\left(1+\epsilon_0\right)\left(\frac{q_1\overline{X}(1+\epsilon_2) + (1-q_1)\overline{X}}{\overline{X}}\right)^{\alpha} \\
{P_1}^* &= \overline{Y}\left(1+\epsilon_0\right)\left(1+q_1\epsilon_2\right)^{\alpha} \qquad (4.1)
\end{aligned}
$$

Assuming $|\epsilon_2| < 1$, the binomial expansion of $(1+q_1\epsilon_2)^{\alpha}$ is applied and terms upto the second order of approximation have been retained. Thus, we obtain

$$
{P_1}^* = \overline{Y}\left(1 + \epsilon_0 + q_1\alpha\epsilon_2 + q_1\alpha\epsilon_0\epsilon_2 + \frac{\alpha(\alpha-1)}{2}q_1^2\epsilon_2^2\right) \qquad (4.2)
$$

Subtracting $\overline{Y}$ from the expression (4.2), we get

$$
{P_1}^* - \overline{Y} = \overline{Y}\left(\epsilon_0 + q_1\alpha\epsilon_2 + q_1\alpha\epsilon_o\epsilon_2 + \frac{\alpha(\alpha-1)}{2}q_1^2\epsilon_2^2\right) \qquad (4.3)
$$

Taking expectation on both sides of equation (4.3) and using the assumptions $E(\epsilon_0) = E(\epsilon_2) = 0$, along with the known moments of the error terms given in Section 2, we get Bias($P_1{}^*$) as:

$$\text{Bias}(P_1{}^*) = E(P_1{}^* - \overline{Y}) = \overline{Y} f_n \left(\frac{\alpha(\alpha - 1)}{2} q_1^2 C_x^2 + q_1 \alpha \rho C_y C_x \right). \tag{4.4}$$

Squaring equation (4.3), neglecting terms of order higher than two, and taking expectation, the mean square error (MSE) of the estimator $P_1{}^*$ to the 1^{st} order of approximation has been obtained as:

$$\text{MSE}(P_1{}^*) = E(P_1{}^* - \overline{Y})^2 = \overline{Y}^2 \left[f_r C_y^2 + f_n \left(q_1^2 \alpha^2 C_x^2 + 2 q_1 \alpha \rho C_y C_x \right) \right]. \tag{4.5}$$

Partially differentiating MSE($P_1{}^*$) with respect to (w.r.t.) α and equating it to zero, we get $\alpha_{opt.}$ as:

$$\alpha_{opt.} = -\frac{\rho C_y}{C_x} \frac{1}{q_1}. \tag{4.6}$$

Now, putting $\alpha_{opt.}$ in MSE($P_1{}^*$), we acquire minimum MSE of $P_1{}^*$ as:

$$\text{min. MSE}(P_1{}^*) = \overline{Y}^2 \left(f_r C_y^2 - f_n \rho^2 C_y^2 \right). \tag{4.7}$$

In the similar manner, considering the developed generalized class of estimator $P_2{}^* = \overline{y}_r \left(\frac{q_2 \overline{x}_r + (1 - q_2) \overline{X}}{\overline{X}} \right)^\beta$.

Making use of the notations specified in section **2**, the estimator previously indicated has been expressed as:

$$P_2{}^* = \overline{Y} (1 + \epsilon_0) \left(\frac{q_2 \overline{X} (1 + \epsilon_1) + (1 - q_2) \overline{X}}{\overline{X}} \right)^\beta$$

$$P_2{}^* = \overline{Y} (1 + \epsilon_0) (1 + q_2 \epsilon_1)^\beta \tag{4.8}$$

Assuming $|\epsilon_1| < 1$, the binomial expansion of $(1 + q_2 \epsilon_1)^\beta$ is applied and terms upto the second order of approximation have been retained. Thus, we obtain

$$P_2{}^* = \overline{Y} \left(1 + \epsilon_0 + q_2 \beta \epsilon_1 + q_2 \beta \epsilon_0 \epsilon_1 + \frac{\beta(\beta - 1)}{2} q_2^2 \epsilon_1^2 \right) \tag{4.9}$$

Subtracting $\overline{Y}$ from the expression (4.9), we get

$$P_2{}^* - \overline{Y} = \overline{Y} \left(\epsilon_0 + q_2 \beta \epsilon_1 + q_2 \beta \epsilon_o \epsilon_1 + \frac{\beta(\beta - 1)}{2} q_2^2 \epsilon_1^2 \right) \tag{4.10}$$

Taking expectation on both sides of equation (4.10) and using the assumptions $E(\epsilon_0) = E(\epsilon_1) = 0$, along with the known moments of the error terms given in Section 2, we get Bias($P_2{}^*$) as:

$$\text{Bias}(P_2{}^*) = E(P_2{}^* - \overline{Y}) = \overline{Y} f_r \left(\frac{\beta(\beta - 1)}{2} q_2^2 C_x^2 + q_2 \beta \rho C_y C_x \right). \tag{4.11}$$

Squaring equation (4.10), neglecting terms of order higher than two, and taking expectation, the mean square error (MSE) of the estimator ${P_2}^*$ to the 1^{st} order of approximation has been obtained as:

$$\text{MSE}({P_2}^*) = E({P_2}^* - \overline{Y})^2 = \overline{Y}^2 f_r \left(C_y^2 + q_2^2\beta^2 C_x^2 + 2q_2\beta\rho C_y C_x\right). \tag{4.12}$$

Partially differentiating MSE(${P_2}^*$) w.r.t. β and equating it to zero, we get $\beta_{opt.}$ as:

$$\beta_{opt.} = -\frac{\rho C_y}{C_x}\frac{1}{q_2}. \tag{4.13}$$

Now, putting $\beta_{opt.}$ in MSE(${P_2}^*$), we acquire minimum MSE of ${P_2}^*$ as:

$$\text{min. MSE}({P_2}^*) = \overline{Y}^2 f_r(1-\rho^2)C_y^2. \tag{4.14}$$

The values of α and β depend on ρ, C_y, C_x, q_1 and q_2. And all these values are assumed to be known in the present work. However, when population parameters are unknown, then these values are estimated on the basis of sample observations without having any loss in efficiency of the developed estimators.

5 Generalized class of estimators' particular cases

Particular cases of the developed generalized class of estimators have been shown in Table 1 and 2:

Table 1: Particular cases of the developed generalized class of estimator ${P_1}^*$

Estimators	α	q_1
$P_1 = \overline{y}_r$ Given by Lee et al.(1994)	0	q_1
$P_2 = \overline{y}_r\left(\frac{\overline{X}}{\overline{x}_n}\right)$ Given by Ahmed et al.(2006)	-1	1
$P_3 = \overline{y}_r\left(\frac{\overline{X}}{q_1\overline{x}_n+(1-q_1)\overline{X}}\right)$ Given by Ahmed et al.(2006)	-1	q_1

Table 2: Particular cases of the developed generalized class of estimator ${P_2}^*$

Estimators	β	q_2
$P_1 = \overline{y}_r$ Given by Lee et al.(1994)	0	q_2
$P_4 = \overline{y}_r\left(\frac{\overline{X}}{\overline{x}_r}\right)$ Given by Ahmed et al.(2006)	-1	1
$P_5 = \overline{y}_r\left(\frac{\overline{X}}{q_2\overline{x}_r+(1-q_2)\overline{X}}\right)$ Given by Ahmed et al.(2006)	-1	q_2

Mean square errors (MSEs) of the particular cases of the developed generalized class of estimators have been obtained as:

(i) MSE(P_1)=$\overline{Y}^2 f_r C_y^2$,
(ii) MSE(P_2)=$\overline{Y}^2 \left[f_r C_y^2 + f_n \left(C_x^2 - 2\rho C_y C_x\right)\right]$,
(iii) MSE(P_3)=$\overline{Y}^2 \left[f_r C_y^2 + f_n \left(q_1^2 C_x^2 - 2q_1\rho C_y C_x\right)\right]$,
(iv) MSE(P_4)=$\overline{Y}^2 f_r \left(C_y^2 + C_x^2 - 2\rho C_y C_x\right)$,
(v) MSE(P_5)=$\overline{Y}^2 f_r \left(C_y^2 + q_2^2 C_x^2 - 2q_2\rho C_y C_x\right)$.

6 Analysis of Comparisons

Within this section, mean square errors (MSEs) of the developed generalized class of estimators have been compared with its particular cases.

(i) Compared to P_1, ${P_1}^*$ is more proficient if $MSE({P_1}^*) < MSE(P_1)$
$\implies \overline{Y}^2 \left[f_r C_y^2 + f_n \left(q_1^2\alpha^2 C_x^2 + 2q_1\alpha\rho C_y C_x\right)\right] < \overline{Y}^2 f_r C_y^2$
Solving it further, we get
$\implies 2\rho C_y < -q_1\alpha C_x$
$\implies \rho < -\frac{q_1\alpha C_x}{2C_y}$.

(ii) Compared to P_2, ${P_1}^*$ is more proficient if $MSE({P_1}^*) < MSE(P_2)$
$\implies \overline{Y}^2 \left[f_r C_y^2 + f_n \left(q_1^2\alpha^2 C_x^2 + 2q_1\alpha\rho C_y C_x\right)\right] < \overline{Y}^2 \left[f_r C_y^2 + f_n \left(C_x^2 - 2\rho C_y C_x\right)\right]$
Solving it further, we get
$\implies 2\rho C_y < C_x(1 - q_1\alpha)$
$\implies \rho < \frac{C_x(1-q_1\alpha)}{2C_y}$.

(iii) Compared to P_3, ${P_1}^*$ is more proficient if $MSE({P_1}^*) < MSE(P_3)$
$\implies \overline{Y}^2 \left[f_r C_y^2 + f_n \left(q_1^2\alpha^2 C_x^2 + 2q_1\alpha\rho C_y C_x\right)\right] < \overline{Y}^2 \left[f_r C_y^2 + f_n \left(q_1^2 C_x^2 - 2q_1\rho C_y C_x\right)\right]$
Solving it further, we get
$\implies 2\rho C_y < (1-\alpha)q_1 C_x$
$\implies \rho < \frac{(1-\alpha)q_1 C_x}{2C_y}$.

(iv) Compared to P_1, ${P_2}^*$ is more proficient if $MSE({P_2}^*) < MSE(P_1)$
$\implies \overline{Y}^2 f_r \left(C_y^2 + q_2^2\beta^2 C_x^2 + 2q_2\beta\rho C_y C_x\right) < \overline{Y}^2 f_r C_y^2$
Solving it further, we get
$\implies 2\rho C_y < -q_2\beta C_x$
$\implies \rho < \frac{-q_2\beta C_x}{2C_y}$.

(v) Compared to P_4, ${P_2}^*$ is more proficient if $MSE({P_2}^*) < MSE(P_4)$
$\implies \overline{Y}^2 f_r \left(C_y^2 + q_2^2\beta^2 C_x^2 + 2q_2\beta\rho C_y C_x\right) < \overline{Y}^2 f_r \left(C_y^2 + C_x^2 - 2\rho C_y C_x\right)$
Solving it further, we get
$\implies 2\rho C_y < C_x(1 - q_2\beta)$
$\implies \rho < \frac{C_x(1-q_2\beta)}{2C_y}$.

(vi) Compared to P_5, ${P_2}^*$ is more proficient if $MSE({P_2}^*) < MSE(P_5)$
$\implies \overline{Y}^2 f_r \left(C_y^2 + q_2^2\beta^2 C_x^2 + 2q_2\beta\rho C_y C_x\right) < \overline{Y}^2 f_r \left(C_y^2 + q_2^2 C_x^2 - 2q_2\rho C_y C_x\right)$
Solving it further, we get

$\implies 2\rho C_y < (1-\beta)q_2 C_x$
$\implies \rho < \frac{(1-\beta)q_2 C_x}{2C_y}$.

7 Empirical Study

For empirical study, two different datasets have been taken into consideration.
Data set 1 [Source: Sarndal et al. (1992)]
This dataset incorporates information about Sweden's 284 municipalities during the period 1970 to 1980. In this dataset, study variable y denotes real estate values according to 1984 assessment (in millions of kronor) and ancillary variable x denotes number of municipal employees in 1984.
Values of the population parameters regarding dataset 1 are given as follows:
N=284, $\overline{Y} = 3077.53$, $\overline{X} = 1779.07$, $S_y^2 = 22520027$, $S_x^2 = 18089114$, $C_y = 1.54$, $C_x = 2.39$, q_1=0.5, q_2=0.5, $\alpha_{opt.}$=-1.21, $\beta_{opt.}$=-1.21 and $\rho_{yx} = 0.94$.

Data set 2 [Source: Das (1988)]
This dataset consists of 278 villages/wards under Gajole Police Station of Malda district of West Bengal, India. In this dataset, study variable y denotes the number of agricultural laborers in 1971 and ancillary variable x denotes the number of agricultural laborers in 1961.
Values of the population parameters regarding dataset 2 are given as follows:
N=278, $\overline{Y} = 39.06$, $\overline{X} = 25.11$, $S_y^2 = 3163.65$, $S_x^2 = 1634.35$, $C_y = 1.44$, $C_x = 1.61$, q_1=0.5, q_2=0.5, $\alpha_{opt.}$=-1.28, $\beta_{opt.}$=-1.28 and $\rho_{yx} = 0.72$.

The performance of the developed generalized class of estimators and its members have been examined for different amount of missing values as:
Case I: Simple random sampling without replacement (SRSWOR) has been used to take a 20% sample from the population, and 15% of the sample's values have been presumed to be missing.
For dataset 1: N=284, n=57, r=48 and for dataset 2: N=278, n=56, r=48.

Case II: Again simple random sampling without replacement (SRSWOR) has been used to take a 20% sample from the population, and 25% of the sample's values have been presumed to be missing.
For dataset 1: N=284, n=57, r=43 and for dataset 2: N=278, n=56, r=42.

The formula below has been used to determine the Percentage Relative Efficiency(PRE) of the developed generalized class of estimators and its particular cases w.r.t. $P_1 = \overline{y}_r$:
$PRE(P_i) = \frac{MSE(P_1)}{MSE(P_i)} * 100$,
where P_i=$P_1, P_2, P_3, P_4, P_5, P_1^*$ and P_2^*.

Based on Case I and II, table 3 and table 4 present the empirical outcomes of the developed generalized class of estimators and its members for dataset 1 and 2.

Table 3: Estimators' MSE and PRE w.r.t. $P_1 = \bar{y}_r$ (Dataset 1)

Estimators	Case I		Case II	
	MSE	PRE	MSE	PRE
P_1	388602.3	100.00	442979.0	100.00
P_2	229290.6	169.48	283667.4	156.16
P_3	119025.3	326.48	173402.1	255.46
P_4	191919.0	202.48	218774.1	202.48
P_5	55787.55	696.57	63593.85	696.57
${P_1}^*$	**110673.6**	**351.12**	**165050.4**	**268.39**
${P_2}^*$	**45476.71**	**854.51**	**51840.22**	**854.51**

Table 4: Estimators' MSE and PRE w.r.t. $P_1 = \bar{y}_r$ (Dataset 2)

Estimators	Case I		Case II	
	MSE	PRE	MSE	PRE
P_1	54.52	100.00	63.94	100.00
P_2	38.18	142.78	47.60	134.32
P_3	32.26	169.02	41.67	153.43
P_4	34.77	156.78	40.78	156.78
P_5	27.61	197.47	32.38	197.47
${P_1}^*$	**31.07**	**175.46**	**40.49**	**157.91**
${P_2}^*$	**26.18**	**208.26**	**30.70**	**208.26**

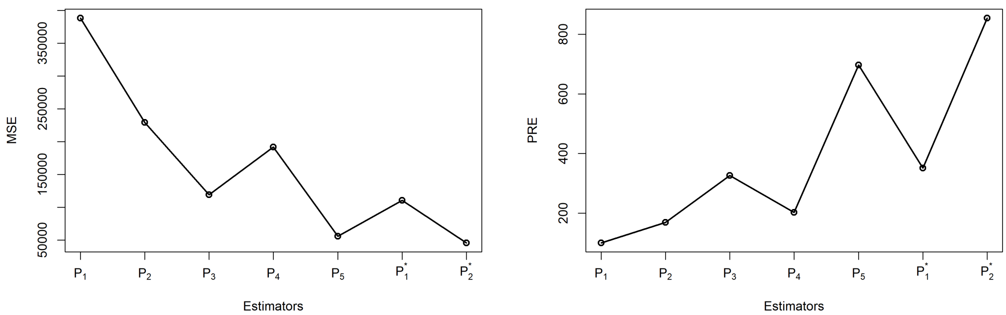

Figure 1: Estimators' MSE and PRE graph for dataset 1 based on Case I

8 Interpretation of the results obtained through empirical study

The proposed generalized class of estimator P_1^* has greater percentage relative efficiency (PRE) than P_1, P_2, and P_3 because to its lower mean square error (MSE), as determined by tables 3 and 4. In the same way, the proposed generalized class of estimator P_2^* has

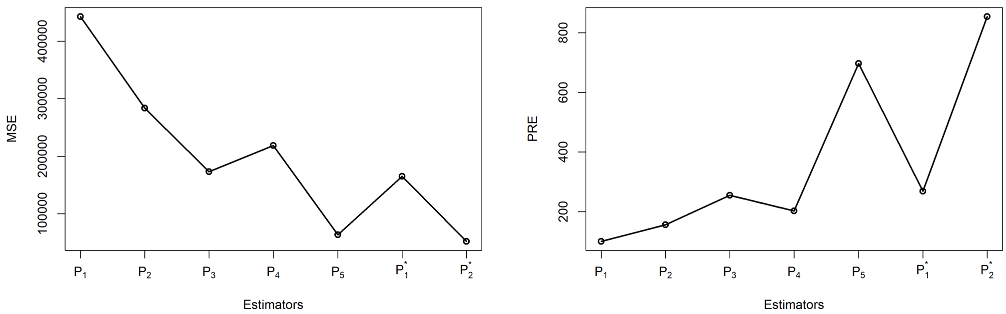

Figure 2: Estimators' MSE and PRE graph for dataset 1 based on Case II

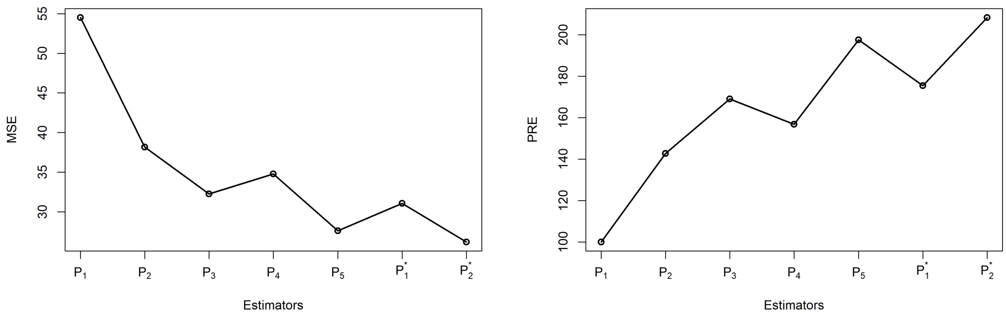

Figure 3: Estimators' MSE and PRE graph for dataset 2 based on Case I

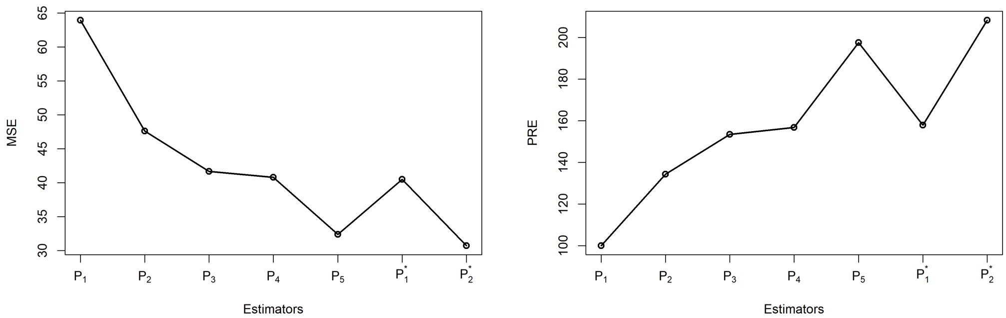

Figure 4: Estimators' MSE and PRE graph for dataset 2 based on Case II

greater percentage relative efficiency (PRE) than P_1, P_4, and P_5 due to its lower mean square error (MSE). Further from figures 1 to 4, we have also obtained that the suggested generalized class of estimator P_1^* has higher percentage relative efficiency (PRE) and lower mean square error (MSE) than the relevant estimators P_1, P_2, and P_3. In a similar way, the proposed generalized class of estimator P_2^* has higher percentage relative efficiency (PRE) than the estimators P_1, P_4, and P_5 due to its lower mean square error (MSE).

9 Conclusion

In the current paper, generalized class of estimators for population mean estimation based on imputation technique under simple random sampling without replacement (SRSWOR) using ancillary variable have been recommended. The mean square error (MSE) and percentage relative efficiency (PRE) results of the recommended estimators have been presented in tables 4 and 5. These results have also been observed through graphs in figures 1, 2, 3 and 4.

Therefore, from tables and figures, it can be concluded that the developed generalized class of estimators has smaller mean square error (MSE) than the traditional mean and ratio method of imputation. The developed generalized class of estimators also outperform estimators given by Ahmed et al. (2006). Hence, it might be advised to the surveyors to further use the discussed general class of estimators as it has greater efficiency than their corresponding estimators mentioned in table 1 and 2.

Declarations

Conflict of interest

The authors declare that there is no conflict of interest.

Authors' Contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Disclaimer (Artificial Intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of this manuscript.

References

Ahmed, M.,S., Al-Titi, O., Al-Rawi, Z., & Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. Statistics in Transition, 7(6), 1247–1264.

Bhushan, S., Kumar, A., Pandey, A.P., & Singh, S. (2023). Estimation of population mean in presence of missing data under simple random sampling. Communications in Statistics-Simulation and computation, 52(12), 6048–6069.

Bhushan, S., & Pandey, A.P. (2018). Optimality of ratio type estimation methods for population mean in the presence of missing data. Communications in Statistics-Theory and Methods, 47(11), 2576–2589.

Das, A.K. (1988). Contribution to the theory of sampling strategies based on auxiliary information (Ph.D.) thesis submitted to Bidhan Chandra Krishi Vishwavidyalaya, Mohanpur, Nadia, West Bengal, India, 1988.

Sinsomboonthong, J., & Sinsomboonthong, S. (2025). New adjusted missing value imputation in multiple regression with simple random sampling and rank set sampling methods. PLOS ONE, 20(3), 1–51.

Lee, H., Rancourt, E., & Sarndal, C.E. (1994). Experiments with variance estimation from survey data with imputed values. Journal of Official Statistics-Stockholm, 10(3), 231–243.

Pandey, M.,K., Singh, G.,N., & Zaman, T. (2025) Estimation of population mean using some improved imputation methods for missing data in sample surveys. Communications in Statistics-Theory and Methods, 54(8), 2378–2392.

Prasad, S. (2018). A study on new methods of ratio exponential type imputation in sample surveys. Hacettepe Journal of Mathematics and Statistics, 47(5), 1281–1301.

Sarndal, C.,E., Swensson, B., & Wretman, J. (1992). Model Assisted Survey Sampling. Springer Series in Statistics.

Singh, S., & Deo, B. (2003). Imputation by power transformation. Statistical Papers, 44, 555–579.

Singh, S., & Horn, S. (2000). Compromised imputation in survey sampling. Metrika, 51, 267–276.