Estimation of Population Mean Using Exponential Type Imputation Technique for Missing Observations

Rajesh Singh
Banaras Hindu University, rsinghstat@gmail.com

Hemant K. Verma
Banaras Hindu University, coolhemant010189@gmail.com

Prayas Sharma
University of Petroleum and Energy Studies, prayassharma02@gmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Estimation of Population Mean Using Exponential Type Imputation Technique for Missing Observations

Rajesh Singh
Banaras Hindu University
Varanasi, India

Hemant K. Verma
Banaras Hindu University
Varanasi, India

Prayas Sharma
University of Petroleum and
Energy Studies
Dehradun, India

Some imputation techniques are suggested for estimating the population mean when the data values are missing completely at random under a simple random sample without replacement scheme. Two classes of point estimators are proposed. The bias and mean squared error expressions of the proposed point estimators are derived up to first order of approximation. It has been shown that the proposed point estimators are more efficient than some existing point estimators due to Lee, Rancourt, and Sarndal (1994) and Singh and Horn (2000). Theoretical findings are supported by an empirical study based on five populations to show the superiority of the constructed estimators and methods of imputation over others.

Keywords: Missing data, imputation, bias, mean squared error, simple random sampling without replacement

Introduction

Missing data is a common and serious problem in survey sampling. Missing data naturally occurs in sample surveys when a few sampling units refuse to respond or are unable to participate in the survey. There are two types of non-responses which occur in surveys: unit non-response and item non-response. Unit non-response occurs when an eligible sample unit fails to participate in a survey because of failure to establish a contact or explicit refusal to cooperate. Item non-response occurs instead when a responding unit does not provide useful answers to particular items

of the questionnaire. Such situations create missing data problem. The imputation is a well-defined methodology by virtue of which such problems can be unraveled.

In the literature several imputation techniques are available and discussed. Rubin (1976) addressed three concepts: observed at random (OAR), missing at random (MAR), and parametric distribution (PD). Rubin defined MAR as the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the value of the unobserved data. Heitjan and Basu (1996) distinguished the meaning of MAR and missing completely at random (MCAR) in a very nice way. The imputation technique is also applicable when information on auxiliary variable is available. Lee et al. (1994; 1995) used the information on an auxiliary variable for the purpose of imputation, Singh and Horn (2000) suggested a compromised method of imputation, Ahmed, Al-Titi, Al-Rawi, and Abu-Dayyeh (2006) suggested several new imputation based estimators that use the information on an auxiliary variable and compared their performances with the mean method of imputation, and Rao and Sitter (1995) used the imputation techniques for variance estimation under two phase sampling. Kadilar and Cingi (2008) and Diana and Perri (2010) also suggested some imputation techniques in case of missing data. In the present study we implicitly assume MCAR.

Let

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$

be the population mean of study variable Y. A simple random sample without replacement (SRSWOR), s, of size n is drawn from $\Omega = \{1, 2,\ldots, N\}$ to estimate the population mean $\bar{Y}$. Let r be the number of responding units out of sampled n, then the number of non-responding units is $(n - r)$. Let the set of responding units be denoted by R and that of non-responding units be denoted by R^c. For every unit $i \in R$, the value y_i is observed. However for the units $i \in R^c$, the y_i values are missing and imputed values are to be derived. We assume that imputation is carried out with the aid of a quantitative auxiliary variable x such that, the value of x for unit i is x_i, known and positive for every $i \in s$. In other words, the data $x_s = \{x_i : i \in s\}$ are known.

Some Available Methods of Imputation and Estimators

There are some classical methods of imputation which are commonly used and given as follows:

Mean Method of Imputation

In this method of imputation, the study variable y after imputation takes the form as

$$y_{\cdot i} = \begin{cases} y_i, & i \in R \\ \bar{y}_r, & i \in R^c \end{cases} \tag{1}$$

and the point estimator of the population mean $\bar{Y}$ is given by

$$\bar{y}_s = \frac{1}{n} \sum_{i \in s} y_{\cdot i} \ . \tag{2}$$

Thus, under this method of imputation, the point estimator of the population mean $\bar{Y}$ is

$$\bar{y}_m = \frac{1}{r} \sum_{i \in R} y_i = \bar{y}_r \ . \tag{3}$$

Lemma 1. The expression of Bias and Variance of the point estimator $\bar{y}_m$ is given as

$$\text{Bias}\left(\bar{y}_m\right) = 0 \tag{4}$$

$$V\left(\bar{y}_m\right) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 \tag{5}$$

where $S_y^2 = \dfrac{1}{N-1} \sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2$.

Ratio Method of Imputation

Following the notations of Lee et al. (1994), in the case of single value imputation, if the ith unit requires imputation, the value $\hat{b}x_i$ is imputed. Thus, the study variable y after imputation takes the form as

$$y_{.i} = \begin{cases} y_i, & i \in \mathrm{R} \\ \hat{b}x_i, & i \in \mathrm{R}^c \end{cases} \tag{6}$$

where

$$\hat{b} = \frac{\sum\limits_{i \in \mathrm{R}} y_i}{\sum\limits_{i \in \mathrm{R}} x_i} \ .$$

Under this method of imputation, the point estimator of the population mean $\bar{Y}$ is given by

$$\bar{y}_{\mathrm{RAT}} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \tag{7}$$

where $\bar{x}_n = \dfrac{1}{n}\sum\limits_{i \in s} x_i$, $\bar{x}_r = \dfrac{1}{r}\sum\limits_{i \in \mathrm{R}} x_i$, and $\bar{y}_r = \dfrac{1}{r}\sum\limits_{i \in \mathrm{R}} y_i$.

Lemma 2. The expression of Bias and Mean Square Error (MSE) of the point estimator $\bar{y}_{\mathrm{RAT}}$ is given as

$$\mathrm{Bias}\left(\bar{y}_{\mathrm{RAT}}\right) = \bar{Y}\left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - \rho C_y C_x\right) , \tag{8}$$

$$\mathrm{MSE}\left(\bar{y}_{\mathrm{RAT}}\right) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left(S_y^2 + R_1^2 S_x^2 - 2R_1 S_{xy}\right) , \tag{9}$$

where S_y^2 is defined as above and

$$S_x^2 = \frac{1}{N-1}\sum\limits_{i=1}^{N}\left(X_i - \bar{X}\right)^2, S_{xy} \frac{1}{N-1}\sum\limits_{i=1}^{N}\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)$$

$$R_1 = \frac{\bar{Y}}{\bar{X}}, C_y = \frac{S_y}{\bar{Y}}, C_x = \frac{S_x}{\bar{X}}, \rho = \frac{S_{xy}}{S_x S_y}$$

Compromised Method of Imputation

Singh and Horn (2000) proposed compromised imputation procedure. After imputation the study variable takes form as

$$y_{\bullet i} = \begin{cases} \dfrac{\alpha n y_i}{r + (1-\alpha)\hat{b}x_i}, & i \in \mathrm{R} \\ (1-\alpha)\hat{b}x_i, & i \in \mathrm{R}^c \end{cases} \tag{10}$$

where α is a suitably chosen constant such that the variance of the resultant estimator is minimum. Here, we are also using information from imputed values for the responding units in addition to non-responding units.

Thus, under compromised method of imputation, the point estimator of the population mean $\bar{Y}$ is

$$\bar{y}_{\mathrm{COMP}} = \alpha \bar{y}_r + (1-\alpha)\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} . \tag{11}$$

Lemma 3. The expression of Bias and MSE of the point estimator $\bar{y}_{\mathrm{COMP}}$ is given as

$$\mathrm{Bias}(\bar{y}_{\mathrm{COMP}}) = \bar{Y}(1-\alpha)\left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - \rho C_y C_x\right) \tag{12}$$

$$\mathrm{MSE}(\bar{y}_{\mathrm{COMP}}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left(S_y^2 + R_1^2 S_x^2 - 2R_1 S_{xy}\right)$$
$$-\left(\frac{1}{r} - \frac{1}{n}\right)\bar{Y}^2\alpha^2 C_x^2 \tag{13}$$

where $\alpha_{\mathrm{opt}} = 1 - \rho\dfrac{C_y}{C_x}$. Thus

$$\mathrm{MSE}(\bar{y}_{\mathrm{COMP}})_{\min} = \left[\left(\frac{1}{n} - \frac{1}{N}\right) - \left(\frac{1}{r} - \frac{1}{n}\right)\rho^2\right]S_y^2 \tag{14}$$

362

Along similar lines, Ahmed et al. (2006) proposed several new imputation techniques by introducing some unknown parameters and hence proposed the corresponding estimators for estimating the finite population means $\bar{Y}$.

Proposed Imputation Methods and Corresponding Estimators

The following two imputation methods are suggested. After imputation for the first proposed imputation of technique, the study variable takes the form as

$$
y_{\bullet i} = \begin{cases} y_i, & i \in \mathrm{R} \\ \dfrac{1}{n-r} \bar{y}_r \left[n \exp\left\{ \alpha \dfrac{\bar{X}^{\frac{1}{h}} - \bar{x}_r^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}_r^{\frac{1}{h}}} \right\} - r \right], & i \in \mathrm{R}^c \end{cases} \tag{15}
$$

where a, h, and α are suitably chosen constants. We optimize α in such a way that the MSE of the resultant estimator is minimum. Thus we have the following theorem:

Theorem 1. Under the proposed method of imputation considered in (15), the point estimator of the population mean $\bar{Y}$ is given as

$$
T_P = \bar{y}_r \exp\left\{ \alpha \frac{\bar{X}^{\frac{1}{h}} - \bar{x}_r^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}_r^{\frac{1}{h}}} \right\}. \tag{16}
$$

Proof:

$$
T_P = \frac{1}{n} \sum_{i \in s} y_{\bullet i} = \frac{1}{n} \left[\sum_{i \in \mathrm{R}} y_{\bullet i} + \sum_{i \in \mathrm{R}^c} y_{\bullet i} \right] \tag{17}
$$

where R and R^c are the sets of responding and non-responding units in the sample s of size n.

Now putting the values from (15) into (17), the point estimator of population is obtained as mean $\bar{Y}$ as defined in (16), which completes the proof.

Table 1. Members of the class of estimators T_P

Estimators		Constants	
$\alpha = 1$	$\alpha = 1$	a	h
$T_{P1} = \bar{y}_r \exp\left[\dfrac{\bar{X} - \bar{x}_r}{\bar{X}}\right]$	$T_{P5} = \bar{y}_r \exp\left[\dfrac{\bar{x}_r - \bar{X}}{\bar{X}}\right]$	1	1
$T_{P2} = \bar{y}_r \exp\left[\dfrac{\sqrt{\bar{X}} - \sqrt{\bar{x}_r}}{\sqrt{\bar{X}}}\right]$	$T_{P6} = \bar{y}_r \exp\left[\dfrac{\sqrt{\bar{x}_r} - \sqrt{\bar{X}}}{\sqrt{\bar{X}}}\right]$	1	2
$T_{P3} = \bar{y}_r \exp\left[\dfrac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r}\right]$	$T_{P7} = \bar{y}_r \exp\left[\dfrac{\bar{x}_r - \bar{X}}{\bar{X} + \bar{x}_r}\right]$	2	1
$T_{P4} = \bar{y}_r \exp\left[\dfrac{\sqrt{\bar{X}} - \sqrt{\bar{x}_r}}{\sqrt{\bar{X}} + \sqrt{\bar{x}_r}}\right]$	$T_{P8} = \bar{y}_r \exp\left[\dfrac{\sqrt{\bar{x}_r} - \sqrt{\bar{X}}}{\sqrt{\bar{X}} + \sqrt{\bar{x}_r}}\right]$	2	2

Because the point estimator proposed in (16) after imputing the missing values, belongs to a class of estimators. Some members of the proposed class of point estimator defined in (16) are shown in Table 1 for different choice of a, h, and α.

The study variable after imputation for the second proposed imputation of technique becomes

$$y_{\cdot i} = \begin{cases} y_i, & i \in R \\ \dfrac{1}{n-r}\bar{y}_r\left[n\exp\left\{\alpha \dfrac{\bar{X}^{\frac{1}{h}} - \bar{x}_n^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}_n^{\frac{1}{h}}}\right\} - r\right], & i \in R^c \end{cases} \qquad (18)$$

where a, h, and α are suitably chosen constants. We optimize α in such a way that the MSE of the resultant estimator is minimum. Thus we have the following theorem:

Theorem 2. Under the proposed method of imputation considered in (18), the point estimator of the population mean $\bar{Y}$ is given as

$$\bullet T_g = \bar{y}_r \exp\left\{\alpha \frac{\bar{X}^{\frac{1}{h}} - \bar{X}_n^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{X}_n^{\frac{1}{h}}}\right\} \tag{19}$$

Proof:

$$T_g = \frac{1}{n}\sum_{i\in s} y_{\bullet i} = \frac{1}{n}\left[\sum_{i\in R} y_{\bullet i} + \sum_{i\in R^c} y_{\bullet i}\right] \tag{20}$$

where R and R^c are the sets of responding and non-responding units in the sample, s, of size n.

Putting the values from (18) into (20), we get the form of the point estimator of population mean $\bar{Y}$ as defined in (19), which completes the proof.

Some members of the proposed class of point estimator defined in (19) are shown in Table 2 for different choices of a, h, and α.

Properties of the Estimators T_P and T_g

To obtain the bias and MSE expressions of the estimators to the first degree of approximation, we define

$$e_0 = \frac{\bar{y}_r - \bar{Y}}{\bar{Y}}, e_1 = \frac{\bar{x}_r - \bar{X}}{\bar{X}}, e_2 = \frac{\bar{x}_n - \bar{X}}{\bar{X}}$$

such that $\mathrm{E}(e_i) = 0$; $i = 0, 1, 2$, and

$$\mathrm{E}(e_0^2) = \left(\frac{1}{r} - \frac{1}{N}\right)C_y^2, \mathrm{E}(e_1^2) = \left(\frac{1}{r} - \frac{1}{N}\right)C_x^2, \mathrm{E}(e_0 e_1) = \left(\frac{1}{r} - \frac{1}{N}\right)\rho C_y C_x,$$

$$\mathrm{E}(e_2^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2, \mathrm{E}(e_1 e_2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2, \mathrm{E}(e_0 e_2) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho C_y C_x$$

Using above terminology, the bias and MSE of the proposed estimators are given below.

Table 2. Members of the class of estimators T_g

Estimators		Constants	
$\alpha = 1$	$\alpha = 1$	a	h
$T_{g1} = \overline{y}_r \exp\left[\dfrac{\overline{X} - \overline{x}_n}{\overline{X}}\right]$	$T_{g5} = \overline{y}_r \exp\left[\dfrac{\overline{x}_n - \overline{X}}{\overline{X}}\right]$	1	1
$T_{g2} = \overline{y}_r \exp\left[\dfrac{\sqrt{\overline{X}} - \sqrt{\overline{x}_n}}{\sqrt{\overline{X}}}\right]$	$T_{g6} = \overline{y}_r \exp\left[\dfrac{\sqrt{\overline{x}_n} - \sqrt{\overline{X}}}{\sqrt{\overline{X}}}\right]$	1	2
$T_{g3} = \overline{y}_r \exp\left[\dfrac{\overline{X} - \overline{x}_n}{\overline{X} + \overline{x}_n}\right]$	$T_{g7} = \overline{y}_r \exp\left[\dfrac{\overline{x}_n - \overline{X}}{\overline{X} + \overline{x}_n}\right]$	2	1
$T_{g4} = \overline{y}_r \exp\left[\dfrac{\sqrt{\overline{X}} - \sqrt{\overline{x}_n}}{\sqrt{\overline{X}} + \sqrt{\overline{x}_n}}\right]$	$T_{g8} = \overline{y}_r \exp\left[\dfrac{\sqrt{\overline{x}_n} - \sqrt{\overline{X}}}{\sqrt{\overline{X}} + \sqrt{\overline{x}_n}}\right]$	2	2

Theorem 3. The Bias of the estimator T_P is given by

$$\text{Bias}(T_P) = \left(\frac{1}{r} - \frac{1}{N}\right)\overline{Y}\frac{\alpha}{ah}\left[\frac{1}{ah}\left(\frac{\alpha}{2} + a - 1\right)C_x^2 - \rho C_y C_x\right] \tag{21}$$

and the MSE of the estimator T_P is given by

$$\text{MSE}(T_P) = \left(\frac{1}{r} - \frac{1}{N}\right)\overline{Y}^2\left[C_y^2 + \frac{\alpha^2}{a^2 h^2}C_x^2 - 2\frac{\alpha}{ah}\rho C_y C_x\right], \tag{22}$$

where the optimum value of α is given by

$$\alpha_{\text{opt}} = ah\rho\frac{C_y}{C_x}$$

Proof: Expressing the estimator T_P in terms of the e's, we have

$$T_P = \overline{Y}(1+e_0)\exp\left[\alpha\left\{\frac{1-(1+e_1)^{\frac{1}{h}}}{1+(a-1)(1+e_1)^{\frac{1}{h}}}\right\}\right]$$

$$\approx \bar{Y}\left(1+e_0\right)\exp\left[\frac{-\alpha e_1}{ah}\left(1+\frac{e_1}{h}-\frac{e_1}{ah}\right)^{-1}\right]$$

$$\approx \bar{Y}\left(1+e_0\right)\exp\left(\frac{-\alpha e_1}{ah}\right)\exp\left(\frac{\alpha(a-1)}{a^2h^2}e_1^2\right)$$

$$\approx \bar{Y}\left(1+e_0\right)\left[1-\frac{\alpha e_1}{ah}+\frac{\alpha(a-1)}{a^2h^2}e_1^2+\frac{1}{2}\frac{\alpha}{a^2h^2}e_1^2\right]$$

$$\left(T_P-\bar{Y}\right)\approx\left[e_0-\frac{\alpha}{ah}e_1--+\frac{\alpha(a-1)}{a^2h^2}e_1^2+\frac{1}{2}\frac{\alpha}{a^2h^2}e_1^2\right]$$

Taking expectation on both sides, we get the bias expression of estimator T_P as

$$\text{Bias}\left(T_P\right)=\left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}\frac{\alpha}{ah}\left[\frac{1}{ah}\left(\frac{\alpha}{2}+a-1\right)C_x^2-\rho C_y C_x\right]$$

To find the MSE of the estimator T_P, we have

$$\text{MSE}\left(T_P\right)=\text{E}\left(T_P-\bar{Y}\right)^2$$

$$\approx\bar{Y}^2\,\text{E}\left(e_0-\frac{\alpha}{ah}e_1\right)^2$$

$$\approx\bar{Y}^2\left(\text{E}\left(e_0^2\right)+\frac{\alpha^2}{a^2h^2}\text{E}\left(e_1^2\right)-\frac{\alpha}{ah}\text{E}\left(e_0e_1\right)\right)$$

$$\text{MSE}\left(T_P\right)=\left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}^2\left[C_y^2+\frac{\alpha^2}{a^2h^2}C_x^2-2\frac{\alpha}{ah}\rho C_y C_x\right]$$

Partially differentiating above equation with respect to α and equating to zero, we have

$$\frac{\partial\left(\text{MSE}\left(T_P\right)\right)}{\partial\alpha}=\left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}^2\left[2\frac{\alpha^2}{a^2h^2}C_x^2-2\frac{1}{ah}\rho C_y C_x\right]=0$$

Simplifying the above equation, we get the optimum value of α as

$$\alpha_{opt} = ah\rho \frac{C_y}{C_x}$$

Theorem 4. The Bias of the estimator T_g is given by

$$\text{Bias}\left(T_g\right) = \left(\frac{1}{n} - \frac{1}{N}\right)\bar{Y}\frac{\alpha}{ah}\left[\frac{1}{ah}\left(\frac{\alpha}{2} + a - 1\right)C_x^2 - \rho C_y C_x\right] \quad (23)$$

and the MSE of the estimator T_g is given by

$$\text{MSE}\left(T_g\right) = \bar{Y}^2\left[\left(\frac{1}{r} - \frac{1}{N}\right)C_y^2 + \left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{\alpha^2}{a^2h^2}C_x^2 - 2\frac{\alpha}{ah}\rho C_y C_x\right)\right], \quad (24)$$

where the optimum value of α is given by

$$\alpha_{opt} = ah\rho \frac{C_y}{C_x}$$

Proof: The above theorem can be proved in a similar way to the proof of Theorem 3.

Efficiency Comparison

Estimator T_P is more efficient than estimator $\bar{y}_m$ if

$$V\left(\bar{y}_m\right) - \text{MSE}\left[T_P\left(\min\right)\right] > 0 .$$

But

$$\left(\frac{1}{r} - \frac{1}{N}\right)\bar{Y}^2 C_y^2 - \left(\frac{1}{r} - \frac{1}{N}\right)\bar{Y}^2\left(1 - \rho^2\right)C_y^2 > 0$$

$$\left(\frac{1}{r} - \frac{1}{N}\right)\bar{Y}^2 \rho^2 C_y^2 > 0$$

since $\dfrac{1}{r} > \dfrac{1}{N}$. Therefore, T_P is more efficient than $\bar{y}_m$. Similarly,

$$\mathrm{MSE}\left(\bar{y}_{\mathrm{RAT}}\right) - \mathrm{MSE}\left[T_P\left(\min\right)\right] > 0 \text{ if } \left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - 2\rho C_y C_x\right) + \left(\frac{1}{r} - \frac{1}{N}\right)\rho^2 C_y^2 > 0$$

$$\mathrm{MSE}\left(\bar{y}_{\mathrm{COMP}}\right) - \mathrm{MSE}\left[T_P\left(\min\right)\right] > 0 \text{ if } \left(\frac{1}{r} - \frac{1}{n}\right) < \left(\frac{1}{r} - \frac{1}{N}\right)\rho^2$$

$$\mathrm{MSE}\left[T_g\left(\min\right)\right] - \mathrm{MSE}\left[T_P\left(\min\right)\right] > 0 \text{ if } \frac{1}{r} > \frac{1}{n}$$

Thus, from the above results, we can say that the estimator T_P is more efficient than other estimators.

Empirical Study

Five populations, A, B, C, D, and E, are considered. Population A is the artificial population of size $N = 200$ from Shukla, Thakur, Pathak, and Rajput (2009), population B is from Ahmed et al. (2006), population C is from Dass (1988), population D is from Murthy (1967, p. 228), and population E is from Singh, Singh, and Kumar (1976, p. 126) with parameters as given in Table 3.

Let $n = 40$, $r = 35$ for population A, $n = 200$, $r = 180$ for population B, $n = 80$, $r = 72$ for population C, $n = 23$, $r = 20$ for population D, and $n = 6$, $r = 5$ for population E respectively. Then the bias and MSE of the proposed point estimators are given in Table 4 and Table 5 for populations A, B, C, D and E respectively.

Table 3. Parameters for study populations

Population	N	$\bar{Y}$	$\bar{X}$	S_y^2	S_x^2	ρ	C_y	C_x
A	200	42.485	18.515	199.0598	48.5375	0.865200	0.37630	0.33210
B	8306	253.750	343.316	338006.0000	862017.0000	0.522231	2.70436	2.29116
C	278	39.070	25.110	3199.2400	1660.0200	0.720000	1.44770	1.62260
D	80	5182.640	285.130	3370161.0000	73129.9400	0.920000	0.35420	0.94840
E	17	33.290	40.060	287.8600	458.3500	0.720000	0.50970	0.54990

Table 4. Biases of estimators

Estimators	Populations				
	A	**B**	**C**	**D**	**E**
$\bar{y}_m$	0.0000	0.0000	0.0000	0.0000	0.0000
$\bar{y}_{RAT}$	0.0051	0.5749	0.0511	19.9568	0.1117
$\bar{y}_{COMP}$	0.0039	0.2543	0.0328	6.8572	0.0745
$T_P(min)$	-0.0413	-0.9872	-0.8184	-10.3199	-0.3164
$T_g(min)$	-0.0351	-0.8863	-0.1890	-8.5252	-0.2417

Table 5. MSEs of estimators

Estimators	Populations				
	A	**B**	**C**	**D**	**E**
$\bar{y}_m$	4.6921	1837.1169	32.9258	126381.0375	40.6391
$\bar{y}_{RAT}$	4.2110	1867.2341	31.3361	175668.0261	36.9018
$\bar{y}_{COMP}$	4.1599	1785.9043	30.6224	107777.7488	35.6648
$T_P(min)$	1.1798	1336.0843	15.8571	19412.1274	19.5718
$T_g(min)$	2.8938	1387.2968	18.9807	38015.4161	24.5460

Tables 4 and 5 exhibits the bias and MSE of different point estimators and it has been observed from the tables that the estimators based on auxiliary information are more efficient than the one which does not use the auxiliary information such as $\bar{y}_m$ to overcome the imputation problems. Both the proposed classes of estimators T_P and T_g are more efficient than the estimators, $\bar{y}_m$, $\bar{y}_{RAT}$ and $\bar{y}_{COMP}$, scrupulously, T_P has minimum MSE among all the estimators considered here.

Conclusion

Two imputation techniques are suggested using auxiliary information followed by two class of estimators for estimating the population mean in case of data values are MCAR under a SRSWR scheme. In addition, some new members are also generated from two proposed class of estimators using the suitable values of constants. The minimum biases and mean square errors of the proposed class of estimators were determined up to the first order of approximation. It was established theoretically and empirically that the proposed class of estimator performs best among the other estimators considered, and consequently the

corresponding (first proposed) method of imputation is better than the other existing methods and may be recommended for further use.

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.

Dass, A. K. (1988). *Contributions to the theory of sampling strategies based on auxiliary information* (Unpublished doctoral dissertation). Bidhan Chandra Agricultural University, West Bengal, India.

Diana, G., & Perri, P. F. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics – Theory and Methods, 39*(18), 3245-3251. doi: 10.1080/03610920903009400

Heitjan, D. F., & Basu, S. (1996). Distinguishing "missing at random" and "missing completely at random". *The American Statistician, 50*(3), 207-213. doi: 10.1080/00031305.1996.10474381

Kadilar, C., & Cingi, H. (2008). Estimators for the population mean in the case of missing data. *Communications in Statistics – Theory and Methods, 37*(14), 2226-2236. doi: 10.1080/03610920701855020

Lee, H., Rancourt, E., & Sarndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. Journal of Official Statistics, 10(3), 231-243. Retrieved from: http://www.jos.nu/Articles/abstract.asp?article=103231

Lee, H., Rancourt, E., & Sarndal, C. E. (1995). Variance estimation in the presence of imputed data for the generalized estimation system. In *Proceedings of the Survey Research Methods Section*, American Statistical Association, 384-389. Retrieved from: http://www.amstat.org/sections/srms/Proceedings/

Murthy, M. N. (1967). *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Rao, J. N. K., & Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrica, 82*(2), p. 453-460. doi: 10.1093/biomet/82.2.453

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-593. doi: 10.1093/biomet/63.3.581

Shukla, D., Thakur, N. S., Pathak, S., & Rajput, D. S. (2009). Estimation of mean under imputation of missing data using factor-type estimator in two-phase sampling, *Statistics in Transition, 10*(3), 397-414.

Singh, D., Singh, P., & Kumar, P. (1976). Handbook on Sampling Methods. New Delhi, India: Indian Agricultural Statistics Research Institute (ICAR).

Singh, S., & Horn, S. (2000). Compromised imputation in survey sampling. *Metrika, 51*(3), 266-276. doi: 10.1007/s001840000054