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](http://digitalcommons.wayne.edu/jmasm)

**Recommended Citation**

Available at: [http://digitalcommons.wayne.edu/jmasm/vol15/iss1/19](http://digitalcommons.wayne.edu/jmasm/vol15/iss1/19)

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.
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

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

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

$$
\bar{y}_s = \frac{1}{n} \sum_{i=1}^{n} y_{ui}.
$$

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

$$
\bar{y}_m = \frac{1}{r} \sum_{i=1}^{r} y_i = \bar{y}_r.
$$

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

$$
\text{Bias}(\bar{y}_m) = 0
$$

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

where $S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$.

Ratio Method of Imputation

Following the notations of Lee et al. (1994), in the case of single value imputation, if the $i^{th}$ unit requires imputation, the value $\hat{b}x_i$ is imputed. Thus, the study variable $y$ after imputation takes the form as
\[ y_i = \begin{cases} y_i, & i \in R \\ \hat{b}x_i, & i \in R^c \end{cases} \]  

(6)

where

\[ \hat{b} = \frac{\sum_{i \in R} y_i}{\sum_{i \in R} x_i}. \]

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

\[ \bar{Y}_{RAT} = \frac{\bar{y}_n}{\bar{x}_r}, \]

(7)

where \( \bar{x}_n = \frac{1}{n} \sum_{i \in x} x_i \), \( \bar{x}_r = \frac{1}{r} \sum_{i \in R} x_i \), and \( \bar{y}_r = \frac{1}{r} \sum_{i \in R} y_i \).

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

\[
\text{Bias}(\bar{Y}_{RAT}) = \bar{Y} \left( \frac{1}{r} - \frac{1}{n} \right) \left( C_x^2 - \rho C_y C_x \right),
\]

(8)

\[
\text{MSE}(\bar{Y}_{RAT}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{r} - \frac{1}{n} \right) \left( S_y^2 + R_i^2 S_x^2 - 2 R_i S_{xy} \right),
\]

(9)

where \( S_y^2 \) is defined as above and

\[
S_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2, \quad S_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X})
\]

\[
R_i = \frac{\bar{Y}}{\bar{X}}, \quad C_y = \frac{S_y}{\bar{Y}}, \quad C_x = \frac{S_x}{\bar{X}}, \quad \rho = \frac{S_{xy}}{S_x S_y}
\]
Compromised Method of Imputation

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

\[
y_{si} = \begin{cases} 
\frac{\alpha y_i}{r + (1 - \alpha) bx_i}, & i \in R \\
(1 - \alpha) bx_i, & i \in R^c
\end{cases}
\]

where \( \alpha \) 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

\[
\bar{Y}_{\text{COMP}} = \alpha \bar{y}_r + (1 - \alpha) \frac{\bar{x}_u}{\bar{x}_r}.
\]

**Lemma 3.** The expression of Bias and MSE of the point estimator \( \bar{Y}_{\text{COMP}} \) is given as

\[
\text{Bias}(\bar{Y}_{\text{COMP}}) = \bar{Y}(1 - \alpha) \left( \frac{1}{r} - \frac{1}{n} \right) \left( C_x^2 - \rho C_y C_s \right)
\]

\[
\text{MSE}(\bar{Y}_{\text{COMP}}) = \left( \frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left( \frac{1}{r} - \frac{1}{n} \right) \left( S_x^2 + R_x^2 S_y^2 - 2 R_x S_{xy} \right) - \left( \frac{1}{r} - \frac{1}{n} \right) \bar{Y}^2 \alpha^2 C_x^2
\]

where \( \alpha_{\text{opt}} = 1 - \rho \frac{C_s}{C_x} \). Thus

\[
\text{MSE}(\bar{Y}_{\text{COMP}})_{\text{min}} = \left[ \left( \frac{1}{n} - \frac{1}{N} \right) - \left( \frac{1}{r} - \frac{1}{n} \right) \rho^2 \right] S_y^2
\]
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 $\overline{Y}$.

**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

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

where $a$, $h$, and $\alpha$ are suitably chosen constants. We optimize $\alpha$ 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 $\overline{Y}$ is given as

$$T_p = \overline{y}_r \exp \left\{ \alpha \frac{\overline{X}^h - \overline{X}_r}{\overline{X}^h + (a-1)\overline{X}_r} \right\}. \quad (16)$$

**Proof:**

$$T_p = \frac{1}{n} \sum_{i \in R} y_i = \frac{1}{n} \left[ \sum_{i \in R} y_i + \sum_{i \in R^c} y_i \right]$$ (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 $\overline{Y}$ as defined in (16), which completes the proof.
EXPONENTIAL TYPE IMPUTATION TECHNIQUE

Table 1. Members of the class of estimators $T_p$

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Constants</th>
<th>$a$</th>
<th>$h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{p1} = \bar{y}_r \exp \left[ \frac{\bar{X} - \bar{x}_r}{\bar{X}} \right]$</td>
<td>$T_{p5} = \bar{y}_r \exp \left[ \frac{\bar{x}_r - \bar{X}}{\bar{X}} \right]$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$T_{p2} = \bar{y}_r \exp \left[ \frac{\sqrt{X} - \sqrt{x}_r}{\sqrt{X}} \right]$</td>
<td>$T_{p6} = \bar{y}_r \exp \left[ \frac{\sqrt{x}_r - \sqrt{X}}{\sqrt{X}} \right]$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$T_{p3} = \bar{y}_r \exp \left[ \frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r} \right]$</td>
<td>$T_{p7} = \bar{y}_r \exp \left[ \frac{-\bar{x}_r - \bar{X}}{\bar{X} + \bar{x}_r} \right]$</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$T_{p4} = \bar{y}_r \exp \left[ \frac{\sqrt{X} - \sqrt{x}_r}{\sqrt{X} + \sqrt{x}_r} \right]$</td>
<td>$T_{p8} = \bar{y}_r \exp \left[ \frac{-\sqrt{x}_r - \sqrt{X}}{\sqrt{X} + \sqrt{x}_r} \right]$</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

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 $a$.

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

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

where $a$, $h$, and $\alpha$ are suitably chosen constants. We optimize $\alpha$ 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
\[ T_g = \bar{y}, \exp \left\{ \alpha - \frac{1}{\bar{X}_{nn}} - \frac{1}{\bar{X}_{nn}^i} \right\} \] (19)

**Proof:**

\[ T_g = \frac{1}{n} \sum_{i \in s} y_i = \frac{1}{n} \left[ \sum_{i \in R} y_i + \sum_{i \in R^c} y_i \right] \] (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 \( \alpha \).

**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

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

such that \( E(e_i) = 0; \ i = 0, 1, 2, \) and

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

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

Using above terminology, the bias and MSE of the proposed estimators are given below.
Table 2. Members of the class of estimators $T_g$

<table>
<thead>
<tr>
<th>Estimators</th>
<th>$a = 1$</th>
<th>$a = 1$</th>
<th>Constants</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{v1} = \bar{y} \exp \left[ \frac{\bar{X} - \bar{x}}{\bar{X}} \right]$</td>
<td>$T_{v5} = \bar{y} \exp \left[ \frac{\bar{X} - \bar{X}}{\bar{X}} \right]$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$T_{v2} = \bar{y} \exp \left[ \frac{\sqrt{\bar{X}} - \sqrt{\bar{x}}}{\sqrt{\bar{X}}} \right]$</td>
<td>$T_{v6} = \bar{y} \exp \left[ \frac{\sqrt{\bar{X}} - \sqrt{\bar{X}}}{\sqrt{\bar{X}}} \right]$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$T_{v3} = \bar{y} \exp \left[ \frac{\bar{X} - \bar{X}}{\bar{X} + \bar{x}} \right]$</td>
<td>$T_{v7} = \bar{y} \exp \left[ \frac{\sqrt{\bar{X}} - \sqrt{\bar{X}}}{\sqrt{\bar{X} + \bar{x}}} \right]$</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>$T_{v4} = \bar{y} \exp \left[ \frac{\sqrt{\bar{X}} - \sqrt{\bar{x}}}{\sqrt{\bar{X} + \bar{x}}} \right]$</td>
<td>$T_{v8} = \bar{y} \exp \left[ \frac{\sqrt{\bar{X}} - \sqrt{\bar{X}}}{\sqrt{\bar{X} + \bar{x}}} \right]$</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

**Theorem 3.** The Bias of the estimator $T_p$ is given by

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

and the MSE of the estimator $T_p$ is given by

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

where the optimum value of $\alpha$ is given by

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

**Proof:** Expressing the estimator $T_p$ in terms of the $e$'s, we have

$$T_p = \bar{Y} (1 + e_0) \exp \left[ \alpha \left( \frac{1 - (1 + e_i)^{\frac{1}{h}}}{1 + (a - 1)(1 + e_i)^{\frac{1}{h}}} \right) \right]$$

366
\[ \approx \bar{y}(1 + e_0) \exp \left[ \frac{-\alpha e_1}{ah} \left( 1 + \frac{e_1}{h} \frac{e_1}{ah} \right)^{-1} \right] \]

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

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

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

Taking expectation on both sides, we get the bias expression of estimator \( T_p \) as

\[
\text{Bias}(T_p) = \left( \frac{1}{r} - \frac{1}{N} \right) \bar{y} \frac{\alpha}{ah} \left[ \frac{1}{ah} \left( \frac{\alpha}{2} + a - 1 \right) C_y - \rho C \right]
\]

To find the MSE of the estimator \( T_p \), we have

\[
\text{MSE}(T_p) = \mathbb{E}(T_p - \bar{y})^2
\]

\[
\approx \bar{y}^2 \mathbb{E} \left( e_0 - \frac{\alpha}{ah} e_1 \right)^2
\]

\[
\approx \bar{y}^2 \left( \mathbb{E}(e_0^2) + \frac{\alpha^2}{a^2h^2} \mathbb{E}(e_1^2) - \frac{\alpha}{ah} \mathbb{E}(e_0 e_1) \right)
\]

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

Partially differentiating above equation with respect to \( \alpha \) and equating to zero, we have

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

Simplifying the above equation, we get the optimum value of \( \alpha \) as
EXPONENTIAL TYPE IMPUTATION TECHNIQUE

\[ \alpha_{opt} = ah \rho \frac{C_y}{C_x} \]

**Theorem 4.** The Bias of the estimator \( T_g \) is given by

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

and the MSE of the estimator \( T_g \) is given by

\[
\text{MSE}(T_g) = F^2 \left[ \left( \frac{1}{r} - \frac{1}{N} \right) C_y^2 + \left( \frac{1}{n} - \frac{1}{N} \right) \left( \frac{\alpha^2}{\alpha^2 h^2} C_x^2 - 2 \frac{\alpha}{ah} \rho C_y C_x \right) \right], \hspace{1cm} (24)
\]

where the optimum value of \( \alpha \) 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

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

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 $\frac{1}{r} > \frac{1}{N}$. Therefore, $T_p$ is more efficient than $\overline{y}_m$.

Similarly,

\begin{align*}
\text{MSE}(\overline{y}_{\text{RAT}}) - \text{MSE}[T_p (\text{min})] &> 0 \text{ if } \left( \frac{1}{r} - \frac{1}{n} \right) \left( C_x^2 - 2 \rho C_x C_y \right) + \left( \frac{1}{r} - \frac{1}{N} \right) \rho^2 C_y^2 > 0 \\
\text{MSE}(\overline{y}_{\text{COMP}}) - \text{MSE}[T_p (\text{min})] &> 0 \text{ if } \left( \frac{1}{r} - \frac{1}{n} \right) < \left( \frac{1}{r} - \frac{1}{N} \right) \rho^2 \\
\text{MSE}[T_s (\text{min})] - \text{MSE}[T_p (\text{min})] &> 0 \text{ if } \frac{1}{r} > \frac{1}{n} 
\end{align*}

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**

<table>
<thead>
<tr>
<th>Population</th>
<th>$N$</th>
<th>$\overline{y}$</th>
<th>$\overline{x}$</th>
<th>$S^2_x$</th>
<th>$S^2_y$</th>
<th>$P$</th>
<th>$C_r$</th>
<th>$C_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>200</td>
<td>42.485</td>
<td>18.515</td>
<td>199.0598</td>
<td>48.5375</td>
<td>0.86520</td>
<td>0.37630</td>
<td>0.33210</td>
</tr>
<tr>
<td>B</td>
<td>8306</td>
<td>253.750</td>
<td>343.316</td>
<td>338006.0000</td>
<td>862017.0000</td>
<td>0.522231</td>
<td>2.70436</td>
<td>2.29116</td>
</tr>
<tr>
<td>C</td>
<td>278</td>
<td>39.070</td>
<td>25.110</td>
<td>3199.2400</td>
<td>1660.0200</td>
<td>0.720000</td>
<td>1.44770</td>
<td>1.62260</td>
</tr>
<tr>
<td>D</td>
<td>80</td>
<td>5182.640</td>
<td>285.130</td>
<td>3370161.0000</td>
<td>73129.9400</td>
<td>0.920000</td>
<td>0.35420</td>
<td>0.94840</td>
</tr>
<tr>
<td>E</td>
<td>17</td>
<td>33.290</td>
<td>40.060</td>
<td>287.8600</td>
<td>458.3500</td>
<td>0.720000</td>
<td>0.50970</td>
<td>0.54990</td>
</tr>
</tbody>
</table>
Table 4. Biases of estimators

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Populations</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{y}_m$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>$\overline{y}_RAT$</td>
<td>0.0051</td>
<td>0.5749</td>
<td>0.0511</td>
<td>19.9568</td>
<td>0.1117</td>
<td></td>
</tr>
<tr>
<td>$\overline{y}_COMP$</td>
<td>0.0039</td>
<td>0.2543</td>
<td>0.0328</td>
<td>6.8572</td>
<td>0.0745</td>
<td></td>
</tr>
<tr>
<td>$T_P$(min)</td>
<td>-0.0413</td>
<td>-0.9872</td>
<td>-0.8184</td>
<td>-10.3199</td>
<td>-0.3164</td>
<td></td>
</tr>
<tr>
<td>$T_d$(min)</td>
<td>-0.0351</td>
<td>-0.8863</td>
<td>-0.1890</td>
<td>-8.5252</td>
<td>-0.2417</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. MSEs of estimators

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Populations</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{y}_m$</td>
<td>4.6921</td>
<td>1837.1169</td>
<td>32.9258</td>
<td>126381.037</td>
<td>40.6391</td>
<td></td>
</tr>
<tr>
<td>$\overline{y}_RAT$</td>
<td>4.2110</td>
<td>1867.2341</td>
<td>31.3361</td>
<td>175668.0261</td>
<td>36.9018</td>
<td></td>
</tr>
<tr>
<td>$\overline{y}_COMP$</td>
<td>4.1599</td>
<td>1785.9043</td>
<td>30.6224</td>
<td>107777.7488</td>
<td>35.6648</td>
<td></td>
</tr>
<tr>
<td>$T_P$(min)</td>
<td>1.1798</td>
<td>1336.0843</td>
<td>15.8571</td>
<td>19412.1274</td>
<td>19.5718</td>
<td></td>
</tr>
<tr>
<td>$T_d$(min)</td>
<td>2.8938</td>
<td>1387.2968</td>
<td>18.9807</td>
<td>38015.4161</td>
<td>24.5460</td>
<td></td>
</tr>
</tbody>
</table>

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 $\overline{y}_m$ to overcome the imputation problems. Both the proposed classes of estimators $T_P$ and $T_R$ are more efficient than the estimators, $\overline{y}_m$, $\overline{y}_RAT$ and $\overline{y}_COMP$, 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


